



**Tendencias de mercado laboral en Colombia mediante la implementación de algoritmos de minería de datos.**

**Carlos Eduardo Camacho Cardenas**

**Institución Universitaria Politécnico Grancolombiano  
Ingeniería de sistemas  
Bogotá D.C.  
2019**

## Resumen

El desempleo en Colombia está en aumento, actualmente, el último trimestre del año arroja una cifra de desempleo un poco mayor a 10%, lo que supone un crecimiento anual con respecto al 2018 de 1.1 puntos porcentuales y una clara tendencia de ascenso en los niveles de personas desempleadas en el país. Este es un tema muy sonado en los principales medios de comunicación del país y uno de los defectos que siempre ha influido en la calidad de vida de las personas que en él residen, Sin embargo, del otro lado, se tienen las brechas que crean la demanda de fuerza laboral de algunos sectores productivos del país, demanda que la misma oferta de graduado y trabajadores no puede suplir, como un ejemplo tenemos el déficit que enfrenta el sector de las TIC en Colombia, se estima que cerca de 45.000 profesionales son necesarios para cumplir con la demanda del sector tecnológico del país ((28-01-2020).Colombia necesita más profesionales en tecnología). El sector de la medicina también se ve afectado, la escasez de especialistas crea una barrera entre la expectativa de los usuarios y el nivel de atención que estos reciben, lo que se traduce en un faltante de 19.000 especialistas para cubrir las necesidades de los usuarios inscritos a servicios de salud en el país.

Basados en esto, el objetivo de este proyecto fue estructurar un sistema de información, fundamentado en el uso de algoritmos de pronóstico de minería de datos que permitió, evidenciar el comportamiento de la oferta y la demanda del mercado laboral colombiano, además de realizar análisis de previsión y necesidad de fuerza laboral en diferentes sectores económicos y regiones del país, junto con la entrada de profesionales a dicho mercado. El análisis de los datos arroja resultados bastante distantes

dependiendo la región de estudio o, aunque esto depende de la perspectiva que se le quiera dar al estudio, se tienen zonas del país con un déficit generalizado de mano de obra, mientras que, en otros, especialmente en las principales ciudades del país, se observan muy bajos índices de graduados con respecto a la demanda laboral que exige su territorio. Estas fluctuaciones de resultados son las que se explicarán más adelante en este documento.

**Palabras Claves:** Minería de datos, Patrones, Mercado laboral, Tendencia, Oferta, Demanda, Sistema de información.

## Índice

1.	Problema .....	8
2.	Objetivos .....	10
2.1.	Objetivo General .....	10
2.2.	Objetivos Específicos .....	10
3.	Justificación .....	11
4.	Marco Teórico .....	13
5.	Metodología .....	18
5.1.	Fase de comprensión del negocio o problema .....	18
5.2.	Fase de comprensión de los datos .....	19
5.3.	Fase de preparación de los datos .....	21
5.3.1.	Tratamiento Demanda laboral 2015-2019 .....	23
5.3.2.	Tratamiento Graduados 2001-2018 .....	25
5.4.	Fase de modelado .....	26
5.5.	Fase de Evaluación .....	30
5.6.	Fase de implementación .....	31
6.	Resultados y Discusión .....	34
7.	Cifras y Proyecciones Generales a nivel nacional .....	35
8.	Tendencias según nivel educativo .....	38
9.	Áreas de conocimiento y sectores productivos .....	44
9.1.	Economía, administración y afines .....	45
9.2.	Ingeniería, arquitectura, urbanismo y afines .....	47
10.	Dinámica general - departamentos .....	50
11.	Conclusiones .....	71
12.	Explorador – Anexo .....	73
13.	Enlace de acceso .....	75
14.	Bibliografía .....	76

## Tabla de imágenes

1. Etapas de procesamiento minería de datos .....	13
2. Formato compatible time-series Forecasting .....	23
3. Estructura primaria demanda 2015-2019 .....	23
4. Estructura primaria graduados 2001-2018 .....	24
5. Estructura final demanda 2015-2019 .....	25
6. Estructura intermedia base de datos graduados .....	26
7. Estructura final Graduados 2001-2018 .....	27
8. Modelo utilizado en el procesamiento de los datos .....	28
9. Preprocesamiento en el modelo .....	29
10. Configuración algoritmo holt-winters .....	30
11. Resultados pronóstico modelo asociado .....	32
12. Dinámica de mercado inicial a nivel nacional .....	36
13. Egresados históricos nacionales según nivel educativo .....	39
14. Ofertas históricas por nivel de formación .....	40
15. Dinámica de mercado por nivel educativo: universitario .....	41
16. Dinámica de mercado por nivel educativo: Tecnología .....	41
17. Dinámica de mercado por nivel educativo: Especialización .....	42
18. Distribución histórica de egresados .....	45
19. Crecimiento en número de graduados - Economía, administración y afines .....	46
20. Crecimiento en la demanda laboral para el sector economía, administración y afines .....	46
21. Crecimiento en número de graduados - Ingeniería, arquitectura, urbanismo y afines .....	48
22. Crecimiento en la demanda laboral para el sector Ingeniería, arquitectura, urbanismo y afines .....	48
23. Dinámica de mercado general - Departamento de Cundinamarca .....	51
24. Dinámica de mercado general - Departamento de Antioquia .....	52

25. Dinámica de mercado general - Departamento de Valle del cauca .....	52
26. Dinámica de mercado general - Departamento de Atlántico .....	53
27. Dinámica de mercado general - Departamento de Bolívar .....	53
28. Dinámica de mercado general - Departamento de Norte de Santander .....	54
29. Dinámica de mercado general - Departamento de Amazonas .....	55
30. Dinámica de mercado general - Departamento de Antioquia .....	55
31. Dinámica de mercado general - Departamento de Arauca .....	56
32. Dinámica de mercado general - Departamento de San andrés y providencia .....	56
33. Dinámica de mercado general - Departamento de Atlántico .....	57
34. Dinámica de mercado general - Departamento de Bogotá D.C. ....	57
35. Dinámica de mercado general - Departamento de Bolívar .....	57
36. Dinámica de mercado general - Departamento de Boyacá.....	58
37. Dinámica de mercado general - Departamento de Caldas .....	59
38. Dinámica de mercado general - Departamento de Caquetá .....	59
39. Dinámica de mercado general - Departamento de Casanare .....	60
40. Dinámica de mercado general - Departamento de Cauca .....	60
41. Dinámica de mercado general - Departamento de Cesar .....	61
42. Dinámica de mercado general - Departamento de Chocó .....	61
43. Dinámica de mercado general - Departamento de Córdoba.....	62
44. Dinámica de mercado general - Departamento de Cundinamarca.....	62
45. Dinámica de mercado general - Departamento de Guainía .....	63
46. Dinámica de mercado general - Departamento de Guaviare .....	63
47. Dinámica de mercado general - Departamento de Huila .....	64
48. Dinámica de mercado general - Departamento de Guajira.....	64
49. Dinámica de mercado general - Departamento de Magdalena.....	65
50. Dinámica de mercado general - Departamento de Meta .....	65
51. Dinámica de mercado general - Departamento de Nariño .....	66
52. Dinámica de mercado general - Departamento de Norte de Santander.....	66
53. Dinámica de mercado general - Departamento de Putumayo.....	67
54. Dinámica de mercado general - Departamento de Quindío.....	67

55. Dinámica de mercado general - Departamento de Risaralda .....	68
56. Dinámica de mercado general - Departamento de Santander.....	68
57. Dinámica de mercado general - Departamento de Sucre.....	69
58. Dinámica de mercado general - Departamento de Tolima.....	69
59. Dinámica de mercado general - Departamento de Valle del Cauca.....	70
60. Dinámica de mercado general - Departamento de Vaupés.....	67
61. Pantalla principal explorador – Proyecciones generales .....	73
62. Pantalla proyecciones por nivel educativo .....	74
63. Pantalla proyecciones por sectores económicos .....	74

## **Problemática**

El mercado laboral del país está atravesando, lo que conocemos hoy como un cambio de paradigma, los cambios en las condiciones de trabajo, el perfilamiento que le dan las compañías a sus ofertas de empleo, la creación de nuevas propuestas de trabajo y sin duda, la más grande, la tecnificación que ha tenido que afrontar la industria para soportar la creciente demanda de bienes y servicios que genera el actual ritmo de crecimiento poblacional del mundo y el desarrollo tecnológico, suponen un replanteamiento de las habilidades y capacidades que necesitan los estudiantes para cubrir eficientemente las exigencias del flujo actual de empleo, además de la generación de un modelo de transformación educativa que apoye directamente la adaptabilidad y el desarrollo de dichas cualidades, esto, con el fin de afianzar y fortalecer el correcto acoplamiento entre la oferta de empleados que produce el sistema educativo superior del país y la demanda laboral de la industria.

La búsqueda de talentos ha cambiado, los avances tecnológicos que se han, y se siguen desarrollando en el mundo están dirigiendo un cambio en la demanda de profesionales en el país y el tipo de formación que estos necesitan. Por un lado tenemos la alteración de la dinámica del comportamiento del mercado laboral, la erradicación de antiguos, y la creación de nuevos empleos propicia en el mercado una tendencia de preferencia y necesidad de ciertas carreras sobre otras, y es aquí donde se hacen las siguientes preguntas: ¿qué tanto le apuestan los jóvenes y las universidades a las profesiones de mayor demanda? La respuesta, por lo que sabemos: no lo suficiente. Según un estudio realizado con datos del ministerio de educación (Snies – Sistema



nacional de información de la educación superior), en el periodo que comprende desde el 2001 hasta el 2017 la cifra de graduados aumentó en más de un 300% con 138.658 en el 2001 y 462.367 para el 2017, cifra positiva, sin embargo, dentro de este crecimiento no se destaca la predilección de los estudiantes por escoger carreras relacionadas con el fenómeno de la revolución laboral. Según datos de la Unión europea se calcula que para el 2020, por ejemplo, la brecha de desabastecimiento de fuerza laboral en la industria tecnológica sea de más de 700.000 personas, en el caso de América Latina de 450.000 y finalmente para Colombia, según información del Ministerio de TICs y estimaciones de la Federación Colombiana de Software llegaría a los 70.000 talentos. Por otro lado sabemos, que el modelo de empleabilidad preestablecido se ha transformado y ha dado entrada al acoplamiento entre el conocimiento general y una formación integral en habilidades blandas: creatividad, adaptabilidad, flexibilidad, análisis etc. como principal factor para la selección de perfiles dentro de una vacante profesional.

## **Objetivos**

### **Objetivo general:**

Desarrollar un sistema de información que utilice métodos de minería de datos para procesar dos agrupaciones generales de datos, la primera: oferta de estudiantes para el mercado profesional, y la segunda: ofertas de trabajo, con el fin de generar un sistema de visualización de información interactivo con el que se puedan realizar reportes de comportamiento en la oferta y la demanda de profesionales en el país.

### **Objetivos Específicos:**

- Recolectar y tratar los datos que ofrecen las fuentes de información elegidas.
- Identificación de los modelos de minería adecuados para procesar los datos y generar los análisis propuestos.
- Estructurar el modelo de base de datos que de soporte al almacenamiento y extracción de los datos para su procesamiento.
- Implementar algoritmos de minería de datos para procesar los conjuntos de datos identificados, para generar modelos predictivos , análisis de comportamiento y tendencias.
- Implementar una herramienta para crear un visor de información interactivo con la que el usuario pueda generar informes enriquecidos

## **Justificación**

Entendiendo ahora el contexto, se tiene a las universidades en medio de la oferta y la demanda laboral del mercado, el punto de formación de profesionales en el que se debe preguntar ¿se están formando profesionales de la forma que exige la demanda laboral en torno a las competencias y áreas que se requieren en el mundo actual?

En torno a esta problemática se ha pensado en el desarrollo de un sistema de información que implementa algoritmos de minería de datos junto con técnicas de estructuración y visualización de datos para procesar los conjuntos de datos que brindan las dos principales fuentes de información que sustentan el proyecto: las bases de datos de los resultados académicos de los estudiantes de carreras profesionales del país y los informes de tendencia laboral que realiza el portal web [empleo.com](http://empleo.com), para estructurar y exponer, mediante el uso de una herramienta de visualización de datos interactiva, reportes y análisis hechos a los patrones en las tendencias de comportamiento en la oferta y la demanda de estudiantes y profesionales en el país.

Esta herramienta se realiza con el fin de determinar cómo se comporta la brecha que divide las necesidades de empleo de los sectores productivos del país y el flujo de salida de profesionales que están brindando las universidades. De este modo el sistema permitirá la construcción de modelos predictivos que muestren de manera clara un panorama en la deficiencia de fuerza laboral que afronta el mercado laboral en Colombia identificando patrones en tres factores importantes: cómo se está

comportando la oferta y la demanda en el mercado laboral, las tendencias de selección profesional y habilidades de los estudiantes de educación superior del país, y finalmente, la evolución en las necesidades y ofertas de trabajo de la industria nacional. Todo con el fin de generar reportes que las entidades educativas puedan utilizar para gestionar estrategias que promuevan el estudio de carreras profesionales que mayor demanda laboral tengan actualmente.

## Marco teórico

### Minería de datos

Como fundamento base para la investigación y la elaboración del proyecto, fue necesario definir el concepto de minería de datos y adentrarse en las herramientas y algoritmos que este ofrece en pro de extraer conocimiento de un conjunto de datos. La minería de datos es el proceso de descubrir información útil automáticamente en grandes repositorios de datos. Las técnicas de minería de datos se implementan para recorrer grandes bases de datos para encontrar patrones novedosos y útiles que de otro modo podrían permanecer desconocidas (Pang Ning, 2006, *Introduction to data mining*). Este proceso consta de una serie de etapas de transformación y preprocesamiento, hasta el post procesamiento y los resultados de minería, como se muestra en la siguiente imagen:

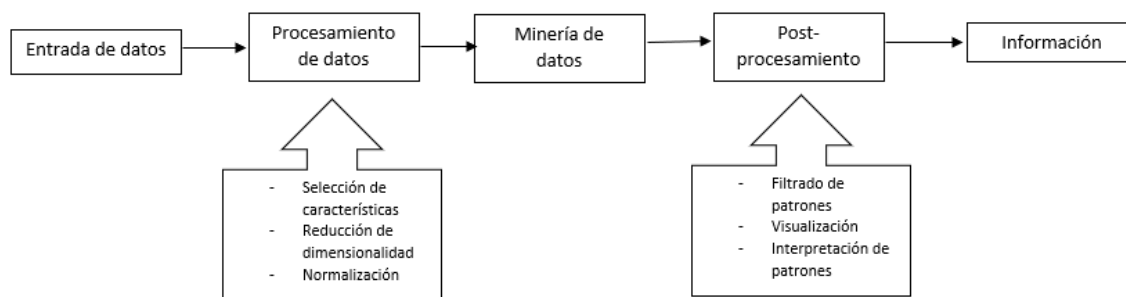


Figura 1 Etapas de procesamiento minería de datos

En la figura anterior se puede evidenciar la segmentación del procesamiento de los datos en diferentes etapas, desde el ingreso de los datos, su preparación mediante

técnicas de preprocesamiento, la aplicación de algoritmos de minería de datos y finalmente los resultados. En el marco de desarrollo del proyecto se tuvo también que aprender acerca del funcionamiento de los distintos algoritmos que la minería de datos ofrece, los cuales están clasificados en algunos grupos principales, como lo son:

- Algoritmos supervisados: que son los que dependen de una variable, en este caso un atributo, sobre el cual se aplicará un modelo de extracción o generación de conocimiento.
- Algoritmos sin supervisar: estos algoritmos son lo opuesto a los ya mencionados supervisados, ya que los modelos que se aplican a los conjuntos de datos no necesitan de un atributo label sobre el cual ejecutar un algoritmo.

De este modo, existen 3 categorías de procesamiento principales: Agrupamiento, Reglas de asociación y predicción. Cada una con un funcionamiento y una resultado distinto dependiendo el objetivo de negocio que su aplicación considere.

### **Algoritmos de pronóstico**

Para efectos del desarrollo de este proyecto, se debió tener en cuenta el estudio de los diferentes algoritmos de pronóstico que se podrían utilizar para realizar las predicciones de los datos. Estos algoritmos nos permiten realizar análisis de series de tiempo, extraer espectros de frecuencias y buscar correlaciones temporales, sobre un o varios atributos objeto de estudio. (Getting started with RapidMiner. <https://docs.rapidminer.com/latest/studio/getting-started/>)

Dentro de los variados repositorios de algoritmos que brinda la minería de datos para realizar los pronósticos que anteriormente se mencionaron, están los siguientes como los principales:

- **Arima:** traducido al español como media móvil integrada autorregresiva, Arima es definido por tres parámetros de *orden*,  $p$ ,  $d$ ,  $q$ , que especifican el número de términos autorregresivos en el modelo.  $d$  especifica el número de diferenciaciones aplicadas a los valores de series de tiempo.  $q$  especifica el número de términos de media móvil en el modelo.

Un modelo ARIMA es un modelo integrado. Describe una serie temporal mediante una suma ponderada de valores de series temporales rezagadas (los términos autorregresivos) y una suma ponderada de residuos rezagados. Estos residuos se originan a partir de un proceso normal de ruido distribuido. El "integrado" indica que los valores del modelo ARIMA están integrados, lo que equivale a que los valores de series temporales originales que describe el modelo están diferenciados.

Este operador es similar a otros operadores de modelado de pronóstico, pero está diseñado específicamente para trabajar en datos de series temporales. Una de las implicaciones de esto es que el modelo de pronóstico debe aplicarse a los mismos datos en los que se capacitó.

- **Holt-Winters:** El método de pronóstico Holt-Winters aplica un suavizado exponencial triple para los componentes de nivel, tendencia y temporada.

Un modelo de Holt-Winters se define por sus tres parámetros de *orden*, alfa, beta, gamma. Alfa especifica el coeficiente para el nivelado de nivel. Beta especifica el coeficiente para el suavizado de tendencias. Gamma especifica el coeficiente para el suavizado estacional. También hay un parámetro para el tipo de estacionalidad: la estacionalidad aditiva, donde cada temporada cambia en un número constante. Estacionalidad multiplicativa, donde cada temporada cambia por un factor. Un factor importante para tener en cuenta es que este operador solo funciona en series de tiempo numéricas.

- **Multi Horizon Forecast:** Un modelo de pronóstico de horizonte múltiple es un metamodelo capaz de predecir múltiples atributos de horizonte a la vez. Para cada atributo de horizonte, el modelo de pronóstico de horizonte múltiple contiene un modelo de predicción. Los atributos del horizonte se seleccionan utilizando una expresión regular en los roles de atributo. La expresión regular por defecto (si el parámetro *expresión regular uso predeterminado* se selecciona) es: "horizonte \ + (\ d +)", que coincide con las funciones de atributos creados por el *Windowing* o de *procesos de Windows* operadores.

El subproceso interno del operador se ejecuta para cada atributo de horizonte seleccionado. Los datos de entrenamiento de entrada, con el atributo de horizonte actual establecido en la función *Etiqueta*, se proporcionan en el puerto del *conjunto de entrenamiento* interno. Se puede usar para entrenar un modelo de predicción para el atributo de horizonte correspondiente. El modelo de predicción



debe proporcionarse al puerto del *modelo* del subproceso interno. Los modelos de predicción se recopilan y se proporcionan como modelo de pronóstico de horizonte múltiple en el puerto de salida del *modelo* del operador. Se pueden pasar objetos adicionales dentro y fuera del subproceso a través de la *entrada y salida* extensor de puerto

### **Baby data**

Otro concepto que se tiene que tomar en consideración fue el de baby data, ya que, en el proceso de desarrollo del proyecto las bases de datos en las que se basaría el procesamiento tuvieron que ser modificadas, en este caso, simplificadas para hacerlas totalmente compatibles con los modelos de minería que escogimos para lograr los resultados. El baby data es considerado un precursor del big data, y a diferencia de éste no se basa en el procesamiento de grandes volúmenes de datos, sino en conjuntos de volumen y formato fácilmente accesibles, informativos y procesables. la idea es simple, lograr una extracción de conocimiento sin tener que incurrir en sistemas de altas prestaciones que usualmente se usan en los análisis de big data. Con small data se hace referencia a el análisis incisivo de pequeños conjuntos de datos extraídos de manera controlada y a escala, dirigidos por una determinada causa relacionada con el objetivo de negocio con el fin de tener datos más accesibles y manejables para convertirlos en conocimiento (Banafa, A. (25-jul-2016). Small Data frente a Big Data: de vuelta a lo básico)

## **Metodología**

Con el fin de optimizar los resultados y estructurar el desarrollo del proyecto, se escogió una metodología orientada a la implementación de proyectos de minería de datos, en este caso, la metodología CRISP (Cross-Industry Standard Process for Data Mining), actualmente es el modelo analítico más usado en el mundo para llevar a cabo proyectos de análisis y estudio de datos, además de ser la única metodología que tiene en cuenta la aplicación al entorno del negocio de los resultados, este modelo ofrece una descripción de las fases que componen el proyecto y las tareas que cada una de estas necesita, además de esto, como un modelo de proceso, ofrece un resumen del ciclo de vida de un proyecto de minería de datos dividiendo, como ya se mencionó antes, el proyecto en 6 fases que se desarrollaron de la siguiente manera:

### **Fase de comprensión del negocio o problema**

En la primera etapa del desarrollo del proyecto se definió el problema como base fundamental del estudio. Se investigó acerca de los fenómenos en el comportamiento de la oferta y la demanda laboral del país y se llegó a la conclusión de que los medios que tenían las personas para acceder a la información que les brindan los datos de mercado estaban de alguna manera u otra, incompletos o fragmentados, con esto se hace referencia a que, en los casos en los que un portal realizaba investigaciones de mercado, este estaba limitado a las ofertas y oferentes que le permitiese su propia base de datos. En otros casos, se evidenció, que los análisis de mercado que se realizaban, estaban muy

segmentados o generalizados, existen actualmente paginas o sistemas de información que muestran estadísticas de los graduados del país, y otros que por su parte, solo exponen metadatos de demanda laboral en Colombia.

Es así como llegamos a la idea de extraer información de ambos sistemas de información, y crear un propio sistema informativo en el que se pueda evidenciar el comportamiento de la oferta y la demanda del mercado laboral en un mismo portal, permitiendo segmentar informes en áreas o períodos específicos de tiempo, para realizar comparaciones y evidenciar cómo se plasman las cifras de desempleo y ocupación en informes gráficos comprensibles.

### **Fase de comprensión de los datos**

La segunda fase, luego de definir el objetivo y la apropiación del proyecto, fue la de la exploración y extracción de datos, estos son la base fundamental de funcionamiento del sistema y el análisis que se pueda hacer con él, así que, se tuvo que explorar varias posibilidades y alternativas, desde informes de estudio de mercados, hasta macro datos nacionales y estadísticas oficiales. En el proceso se descubrieron muchas limitantes, la primera, es que no se tiene una cifra oficial y completa acerca de la oferta de profesionales en el mercado actual, se encontró una base de datos muy rica en información, pero su descarga fue imposible, entonces, se tuvo que adaptar un conjunto diferente: la cifra oficial de graduados a nivel nacional. Por otro lado, se recurrió a la información publicada en el portal de la unidad del servicio público de empleo, entidad adscrita al ministerio de trabajo encargada de realizar estudios e investigaciones del

mercado laboral del país, para extraer los índices de demanda laboral del país a lo largo de un periodo de años, es decir, el número de ofertas de empleo publicadas en los portales adscritos a la entidad de servicio de empleo. Los conjuntos de datos utilizados se especifican a continuación:

1. Demanda Laboral 2015 – 2019:

- Descripción: Cantidad de ofertas laborales desagregadas por año, mes, departamento, municipio, clasificación ocupacional CIUO 08 AC, clasificación sectorial CIIU 4 AC, nivel educativo requerido, experiencia laboral requerida y rangos salariales ofrecidos en las vacantes.
- Sector: Economía - Empleo
- Fuente: <https://serviciodeempleo.gov.co/observatorio/category/demanda-laboral/anexo-estadistico-de-demanda-laboral-nuevo-formato/>

2. Graduados en Colombia 2001-2017:

- Descripción: base de datos que refleja el número total de los graduados del sistema de educación superior colombiano, desagregado por: nivel de formación, sector, sexo, carácter académico de la Institución de Educación Superior, metodología, departamento, área y Núcleo Básico de Conocimiento entre otros.
- Sector: Educación

- Fuente:<https://www.mineducacion.gov.co/sistemasinfo/Informacion-a-la-mano/212400:Estadisticas>

### **Fase de preparación de los datos**


La fase de preparación de los datos se llevó a cabo paralelamente con la de modelado, ya que, de acuerdo con la estructura que necesitaban los datos para ser compatibles con el modelo seleccionado, se transformaban entonces los conjuntos de datos anteriormente especificados. El proceso de transformación y tratamiento se explican a continuación:

1. Estructura general: Las bases de datos que preliminarmente se habían extraído, tenían estructuras diferentes entre sí y el modelo elegido, en nuestro caso Forecast (se explicará su funcionamiento en la siguiente etapa) funciona con estructuras lineales de tiempo, una columna con un atributo fecha, que será la línea de tiempo y otra columna, que será entonces los valores que el algoritmo va a pronosticar de acuerdo a la línea de tiempo que lo define, la estructura se muestra más claramente, con un ejemplo, a continuación:

Row No.	Date	Lake surface level / feet
1	Jun 30, 1875 8:00:00 PM COT	580.380
2	Jun 30, 1876 8:00:00 PM COT	581.860
3	Jun 30, 1877 8:00:00 PM COT	580.970

Figura 2 Formato compatible time-series Forecasting

Y los datos del proyecto, en su estado primitivo tenían las siguientes estructuras:



Código DIVIPOLA	Departamento	2017						
		Mar	Abr	May	Jun	Jul	Ago	Sep
00	Total nacional	129.241	111.005	132.631	119.543	120.032	153.757	151.195
91	Amazonas	110	65	128	133	58	98	94
05	Antioquia	16.243	13.896	17.284	15.332	14.647	19.069	19.027
81	Arauca	269	235	384	357	279	317	418
08	Atlántico	4.228	3.324	4.045	3.689	3.588	4.266	4.112
11	Bogotá D.C.	65.204	56.946	66.419	60.363	61.788	80.626	77.851
13	Bolívar	1.824	1.523	1.882	1.845	1.681	2.252	2.310
15	Boyacá	1.387	1.137	1.543	1.410	1.403	1.844	1.882
17	Caldas	1.056	880	1.096	1.000	1.040	1.283	1.220
18	Caquetá	296	222	249	164	179	235	332
85	Casanare	2.725	1.851	2.268	1.668	1.725	1.715	1.926
19	Cauca	739	538	927	733	634	810	831
20	Cesar	940	651	912	896	870	1.097	985
27	Chocó	165	110	108	140	151	216	190
23	Córdoba	650	599	761	656	542	740	786
25	Cundinamarca	6.718	5.858	6.762	5.752	6.027	7.697	8.069
94	Guainía	29	9	25	7	54	29	31
95	Guaviare	19	12	40	19	43	45	40
41	Huila	1.145	1.015	1.332	1.181	1.192	1.554	1.508
44	La Guajira	350	287	353	332	369	316	427
47	Magdalena	618	690	688	584	682	1.109	902
50	Meta	2.538	2.010	2.556	2.747	2.760	3.329	3.386
52	Nariño	840	733	914	899	819	947	940
54	Norte de Santander	1.038	1.194	1.114	911	891	1.174	1.215
86	Putumayo	347	321	384	523	424	570	551
63	Quindío	589	537	655	640	616	942	853
66	Risaralda	2.040	1.755	2.177	1.739	1.917	2.388	2.346
88	Archipiélago de San Andrés, Providencia y Santa Catalina	225	179	192	147	180	301	277
68	Santander	4.191	3.606	4.365	4.171	3.783	4.902	4.536
70	Sucre	311	269	286	252	318	451	435
72	Tolima	1.558	1.285	1.628	1.317	1.380	1.785	1.677

Figura 3 Estructura primaria demanda 2015-2019

Municipio de oferta del programa	ID Sexo	Sexo	Año*	Semestre	Graduados
BOGOTA DC	2	MUJER	1/12/2011	2	6
BOGOTA DC	2	MUJER	1/12/2009	2	17
BOGOTA DC	2	MUJER	1/12/2012	1	5
BOGOTA DC	2	MUJER	1/12/2012	2	36
BOGOTA DC	2	MUJER	1/12/2011	1	27
BOGOTA DC	1	HOMBRE	1/12/2014	2	44
BOGOTA DC	2	MUJER	1/12/2005	2	28
BOGOTA DC	2	MUJER	1/12/2015	2	15
BOGOTA DC	2	MUJER	1/12/2004	2	5
BOGOTA DC	2	MUJER	1/12/2011	2	15
BOGOTA DC	1	HOMBRE	1/12/2012	1	28
BOGOTA DC	2	MUJER	1/12/2008	1	11
BOGOTA DC	1	HOMBRE	1/12/2009	1	1
BOGOTA DC	1	HOMBRE	1/12/2011	1	7
BOGOTA DC	2	MUJER	1/12/2014	1	5
BOGOTA DC	1	HOMBRE	1/12/2011	2	28
BOGOTA DC	1	HOMBRE	1/12/2004	2	30
BOGOTA DC	1	HOMBRE	1/12/2013	2	7

Figura 4 Estructura primaria graduados 2001-2018

Los datos de cada una de las tablas se trataron de acuerdo con la necesidad de transformación que cada uno tuviese, véase cómo se decidió adaptar ambas bases de datos al modelo:

### Tratamiento Demanda laboral 2015-2019

Para el caso de esta base de datos, la transformación comenzó con eliminar algunas columnas con el fin de simplificar el conjunto de datos, como se puede observar, las ofertas laborales de cada departamento, están desagregadas en años y estos a su vez están divididos en 12 meses, lo que daría un total de 60 atributos, entonces, la manera

más práctica de simplificar el conjunto, sin modificar los datos, y conservar su valor, fue dejar únicamente los atributos que contenían el total de ofertas para cada departamento en todo el año transcurrido.

El siguiente paso, la copia de los datos, se hizo con el fin de cambiar la estructura general del conjunto, es decir, se quería una organización vertical como la que se mostró anteriormente, entonces se copian los valores de cada departamento en un año verticalmente, y se crea una nueva columna [año] y se multiplica verticalmente hacia abajo la lista de departamentos para realizar el proceso anterior la cantidad de años que contenga el conjunto, el proceso derivaría en la siguiente estructura:

Santander	13739	1/12/2015
Sucre	870	1/12/2015
Tolima	4285	1/12/2015
Valle del Cauca	34332	1/12/2015
Vaupés	16	1/12/2015
Vichada	54	1/12/2015
Amazonas	268	1/12/2016
Antioquia	71148	1/12/2016
Arauca	1110	1/12/2016
Atlántico	22222	1/12/2016
Bogotá D.C.	261896	1/12/2016
Bolívar	10617	1/12/2016
Boyacá	6651	1/12/2016
Caldas	5015	1/12/2016
Caquetá	1081	1/12/2016
Casanare	5811	1/12/2016
Cauca	2955	1/12/2016

Fig. 5 estructura final demanda 2015-2019



## Tratamiento Graduados 2001-2018

El tratamiento de este conjunto de datos, por su parte, fue más extenso, gracias a que contiene más de 100 atributos originalmente, se tuvo que primero limpiar eliminando los atributos que serían de importancia nula para el análisis, luego de esto, se quería llegar finalmente a la estructura compatible del modelo, pero como se pudo observar en su estructura original, aparte de ser grande horizontalmente también lo era verticalmente, este es nuestro conjunto de datos más grande, cerca de 465 000 registros, lo que se hizo entonces para transformarla, fue primero, utilizando tablas dinámicas de Excel, darle la forma preliminar de la base de datos de demanda, como se verá a continuación:

Etiquetas de fila	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
AMAZONAS				6	42	10	42	27	28	17	169	139
ANTIOQUIA	16349	17359	18266	19528	19496	21546	22301	25152	28077	29357	40360	47391
ARAUCA	72	99	10	59	29	34	79	30	18	110	343	348
ATLANTICO	8164	7712	8190	7872	7543	5865	11143	10075	10141	10850	13614	15841
BOGOTA DC	56219	51259	52112	54764	58169	67784	65432	81713	84281	88299	109643	114034
BOLIVAR	2434	3461	3326	6224	4446	4184	4689	5318	5335	5264	7362	9764
BOYACA	3052	3396	5208	4215	3605	3593	5215	5215	5025	6971	10039	10903
CALDAS	5297	5339	4524	3967	3573	4024	6056	6582	5371	6345	6767	9078
CAQUETA	191	170	307	368	377	521	992	768	990	1147	1144	1328
CASANARE	98	63	109	123	139	109	200	336	236	374	641	689
CAUCA	908	976	1029	1278	1430	1076	2041	2552	2362	2667	4369	5322
CESAR	1275	954	1222	1265	805	1001	1649	1747	1661	2136	3338	3519
CHOCO	1103	548	805	969	258	694	1672	1382	1456	1375	1945	2356
CORDOBA	2403	1512	1259	2420	1768	2033	2053	2288	2198	2311	2646	3369
CUNDINAMARCA	1959	2755	4075	3661	1785	2270	5588	6667	7027	5931	9452	11556
GUAJIRA	645	752	897	664	463	91	168	1092	1014	1303	1351	1941
GUAVIARE				21			12	36	30	91	98	144

Figura 6 Estructura intermedia base de datos graduados

Luego de esto, el siguiente paso fue simplemente aplicar la misma lógica de transformación de forma vertical como se hizo con la tabla de demanda, lo que daría un resultado final y compatible con el modelo, el resultado fue el siguiente:

Tolima	5258	1/12/2004
Valle del Cauca	12348	1/12/2004
Vaupés	1	1/12/2004
Vichada	9	1/12/2004
Amazonas	42	1/12/2005
Antioquia	19496	1/12/2005
Arauca	29	1/12/2005
Atlántico	7543	1/12/2005
Bogotá D.C.	58169	1/12/2005
Bolívar	4446	1/12/2005
Boyacá	3605	1/12/2005
Caldas	3573	1/12/2005
Caquetá	377	1/12/2005
Casanare	139	1/12/2005

*Figura 7.* Estructura final Graduados 2001-2018

### **Fase de modelado**

El programa utilizado en el proyecto para realizar el modelado fue rapidMiner, entre todas las herramientas de aplicación de minería de datos que se tienen en el mercado, se consideró que es la más amigable e intuitiva, sin restar atención a su completo conjunto de algoritmos y herramientas de preprocesamiento y visualización de resultados.

Dando cumplimiento al objetivo del proyecto se estudiaron varios casos de aplicación de algoritmos de minería de datos a los conjuntos con los que cuenta el proyecto, en esta fase se tienen que tener especial cuidado de escoger modelos que se adaptan también al tipo de datos que se tienen y luego de revisar modelos de clustering, asociación y predicción (los modelos de predicción tienen un funcionamiento distinto a

los de Forecasting), se llegó a la conclusión, de que teniendo un línea de tiempo explícita en los dos conjuntos de datos, se aplicarían modelos de Forecasting o pronóstico, para realizar simplemente, eso, pronósticos que mostraran cómo se comportarían las brechas de desempleo u ocupación, con el fin de generar soluciones que corrigieran el déficit o sobrepoblación de fuerza laboral para algunos sectores productivos del país.

En el marco teórico se expusieron tres algoritmos de pronóstico distintos, holt winters, multivariable, y arima. Para dar cumplimiento al desarrollo de los objetivos del proyecto, se tuvo que decantar por el uso de un algoritmo en concreto, en este caso, se eligió holt winters para llevar a cabo el pronóstico, ¿por que? Recuerde que cada uno funciona de manera distinta, por un lado, multivariable sirve para realizar pronósticos de línea de tiempo sobre múltiples atributos de evaluación, en el caso de los conjuntos del proyecto solo se cuenta con uno, oferta o demanda concretamente. Por otro lado, Arima, ofrece un pronóstico autorregresivo, es decir, que los valores tienden a volver a su valor origen, lo que acota su funcionamiento a casos muy concretos que no se adaptan a las necesidades de este proyecto. Finalmente se tiene Holt-winters, el algoritmo utilizado para pronosticar los datos que fundamentan el objetivo de estudio, este, utiliza un método de suavizado exponencial de tres parámetros (que se explicarán mas adelante) para pronosticar valores a lo largo de una línea de tiempo, teniendo en cuenta las tendencias generales de crecimiento y descenso de los datos a lo largo de la línea plasmada.

Veamos primero el modelo para poder explicar el ¿por qué? y su funcionamiento:

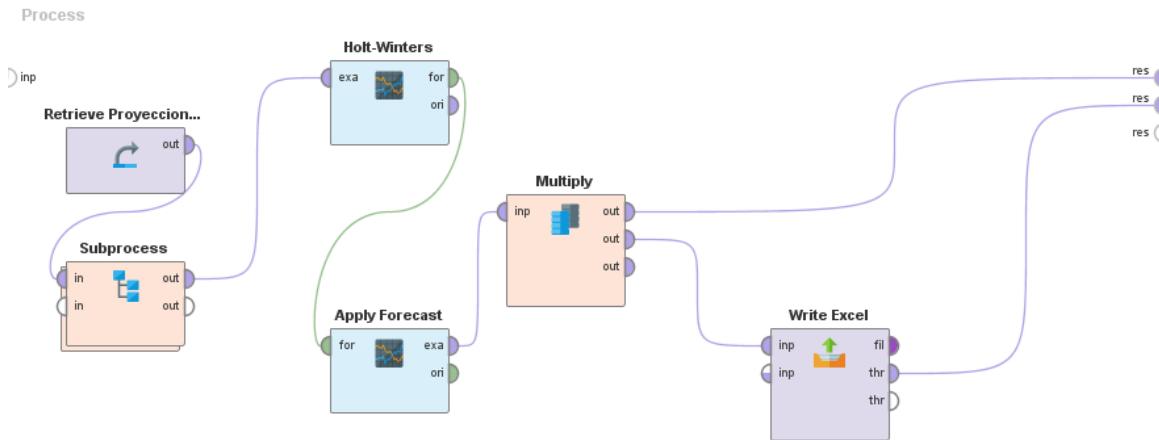


Figura 8. Modelo utilizado en el procesamiento de los datos

**Retrieve:** La primera parte del modelo, recuperación, es simplemente la entrada de datos sobre la cual se va a aplicar el preprocesamiento y el modelo de pronóstico.

**Subprocess:** El bloque de subproceso, encapsula lo que llamamos preprocesamiento de los datos, los módulos elegidos para tratar los datos son los siguientes:

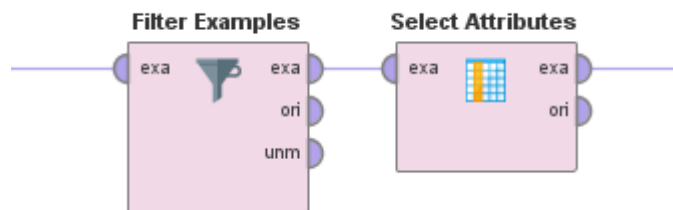


Figura 9. Preprocesamiento en el modelo

Este módulo de preprocesamiento funciona para adaptar y segmentar los conjuntos de datos al modelo de pronóstico. El módulo de filtrar ejemplos sirve para seleccionar, en este caso, un departamento, y sobre este realizar el pronóstico de ofertas y graduados. Este modelo tiene que ser ejecutado unas 32 veces, que es el número de departamentos en el que se divide el territorio nacional. Y finalmente el módulo de seleccionar atributos, limpia la tabla, eliminando columnas innecesarias para ejecutar el algoritmo.

**Holt-winters:** holt winters fue el algoritmo elegido para realizar los pronósticos, en él se aplica un suavizado exponencial triple para los componentes de análisis: error, temporada y horizonte.

Un modelo de Holt-Winters se define por sus tres parámetros de orden: alfa, beta, gamma, que se traducen en los componentes que se han mencionado antes y se explican a continuación:

- Alfa (error) especifica el coeficiente para el nivelado a tener en cuenta con el error que genera una predicción de acuerdo al nivel de horizonte en que esta se encuentre.
- Beta (temporada) especifica el coeficiente para el suavizado de tendencias, en este caso, tiene en cuenta periodos de tiempo en el que el pronóstico se ve influenciado de mayor manera.
- Gamma (horizonte) especifica el coeficiente para el suavizado estacional y en el, se evalúa el horizonte total de pronóstico, para afianzar la proyección y ayudar a disminuir el error promedio de cada predicción.

La configuración que se usó en el módulo fue la siguiente:

Holt-Winters	
time series attribute	Graduados
<input checked="" type="checkbox"/> has indices	
indices attribute	año
alpha: coefficient for level smoothing	0.5
beta: coefficient for trend smoothing	0.1
gamma: coefficient for seasonality smoothing	0.5
period: length of one period	4
seasonality model	multiplicative

*Figura 10.* Configuración algoritmo holt-winters

**Apply Forecast:** Este módulo define el periodo total de tiempo que necesita ser pronosticado, es decir, explicado con un ejemplo, para la base de datos de demanda, que en nuestro caso se extiende en el periodo 2015-2019, se definió un horizonte de 7 años, lo que arrojaría un pronóstico hasta el 2025, que fue el periodo máximo definido dentro del marco del proyecto.

El resto de los módulos se utilizaron como complementos para facilitar la extracción de resultados, migrándolos a hoja de cálculo Excel para ser anexados a las bases de datos originales.

**Fase de evaluación:**

Luego de construir el modelo por completo y transformar las bases de datos con el fin de que se acoplen a este, los primeros resultados fueron las proyecciones que se plantearon al inicio del proyecto como un objetivo dentro de la construcción del sistema de información propuesto, a partir de ellas se pueden obtener las tendencias de oferta y demanda laboral que buscamos analizar en pro de describir las necesidades que tiene el empleo en Colombia y la capacidad que tienen las universidades e instituciones educativas para cubrirlos.

En la fase de evaluación se tuvieron que procesar los dos conjuntos de datos base del proyecto. El modelo arroja proyecciones concretas, es decir para un segmento concreto de información, con lo que cada una, tuvo que reprocesarse un total de 32 veces, para así construir los pronósticos para cada departamento, en el intervalo indicado, como ya mencionamos anteriormente, un horizonte de 7 años, para una línea de tiempo continua desde el 2015 en caso de la base de datos de demanda o desde el 2001 para el conjunto de graduados, hasta juntamente, el año 2025.

Las proyecciones que arroja el modelo tienen la siguiente estructura:

año	forecast of Graduados	Graduados	Graduados and forecast
Dec 1, 2011 ...	?	67	67
Dec 1, 2012 ...	?	78	78
Dec 1, 2013 ...	?	133	133
Dec 1, 2014 ...	?	155	155
Dec 1, 2015 ...	?	132	132
Dec 1, 2016 ...	?	140	140
Dec 1, 2017 ...	?	183	183
Dec 1, 2018 ...	?	185	185
Dec 1, 2019 ...	278.583	?	278.583
Nov 30, 2020 ...	215.925	?	215.925
Nov 30, 2021 ...	180.956	?	180.956
Nov 30, 2022 ...	257.501	?	257.501
Nov 30, 2023 ...	348.627	?	348.627
Nov 29, 2024 ...	267.004	?	267.004
Nov 29, 2025 ...	221.372	?	221.372

*Figura 11.* Resultados pronostico modelo asociado.

La siguiente etapa a la recolección de los pronósticos resultantes de la evaluación del modelo, es la construcción final de las bases de datos, es decir, unir la línea de tiempo actual, períodos 2015 - 2019 con la serie pronóstico, así entonces, los conjuntos de datos cambiarían sus intervalos de valores.

### **Fase de implementación**

La fase final del proyecto consistió en iniciar la construcción del sistema de información propuesto, un observador de tendencias de oferta y demanda en el mercado laboral colombiano. Para dar cumplimiento a este objetivo, se optó por el uso del



programa power-bi, un servicio de análisis empresarial de Microsoft con principal objetivo, proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial con una interfaz lo suficientemente simple para usuarios finales. se eligió esta herramienta, gracias a su facilidad de uso y su completa librería de funcionalidades, lo que la convierte en mucho más que un simple creador de informes, teniendo el conocimiento y la experiencia necesarias, puede llegar a crear sistemas de información complejos, interactivos e intuitivos para cualquier tipo de usuarios.

## **Resultados Análisis y discusión**

A partir de los resultados que arrojaron los modelos predictivos que se usaron para procesar y pronosticar los datos, se pueden plantear ciertos análisis que describen el comportamiento del mercado laboral en Colombia, sus necesidades y sobre todo las tendencias que marcan las profundas necesidades de un mercado en constante evolución como lo es el nuestro.

En este segmento, se centró el análisis concretamente en ver, cómo evolucionan las necesidades de la industria y en su misma medida, hasta qué punto la oferta de profesionales que ofrecen las instituciones educativas superiores de Colombia es capaz de cubrir exitosamente los requerimientos generales de los distintos sectores productivos que conforman la economía colombiana.

Las curvas de predicción que brinda el modelo evaluado en fases anteriores posibilitan seguir con detenimiento la evolución del comercio laboral en los próximos años, con el que podríamos identificar patrones clave que debería seguir la comunidad educativa en su objetivo de adaptarse a la cambiante demanda de profesionales y personal capacitado por la que atraviesa la comunidad actual.

## Cifras y proyecciones generales a nivel nacional

La primera curva que se analizará será la de tendencia general por departamentos, en este segmento se tienen dos pronósticos paralelos, la cantidad de ofertas laborales disponibles en el país a lo largo de cada año dentro del intervalo al cual fue reducido y la oferta de la que dispone el mercado, representada no de forma tradicional, sino en los índices de egresados en programas de post y pregrado a lo largo del país, estos índices muestran de forma indirecta, que tan fundamentado está el mercado en términos de la oferta que lo soporta, desde el punto inicial de su salida activa a la dinámica laboral: la educación.

Las cifras de oferta y demanda laboral para el intervalo de tiempo inicial,

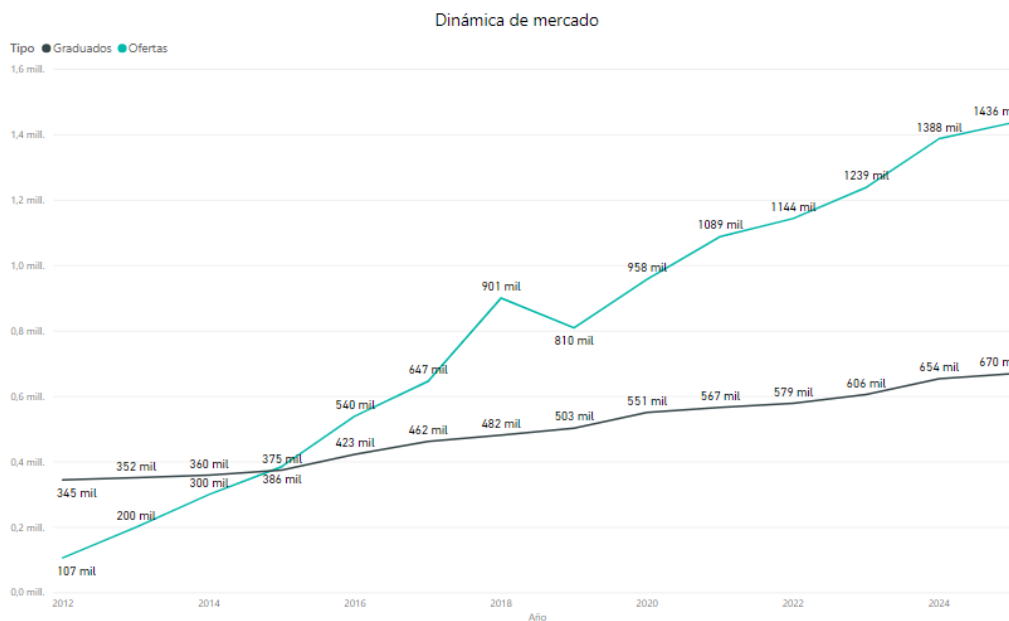


Figura 12. Dinámica de mercado inicial a nivel nacional

La figura anterior describe entonces, la dinámica que ha tenido el mercado a lo largo de los últimos años, en este caso se decidió definir un rango fijo en el que se pudiera observar el comportamiento de la oferta y la demanda de forma paralela entre los dos conjuntos, ya que ambos, originalmente tienen un inicio de intervalo distinto, y aunque el pronóstico que se realizó tiene el mismo horizonte para ambos, el año de inicio de muestreo de los datos, es muy distante, concretamente 14 años, 2001 para la base de datos de graduados y 2015 para la de demanda laboral.

La primera gráfica muestra un patrón claramente identificable, ambas curvas tienen tendencia de ascenso, aunque una de las dos con un grado de crecimiento anual mucho mayor, la curva de ofertas de trabajo para demanda laboral posee un ritmo de crecimiento anual promedio de 24%, un índice de crecimiento demasiado acelerado en comparación con los principales indicadores del país., mientras que del otro lado, la curva de graduados para oferta laboral de primera instancia, tiene una tasa de crecimiento promedio anual de 7.7%, un porcentaje mucho más ajustado al ritmo de crecimiento económico del país.

Los ritmos de crecimiento plasmados en la figura 11 permite realizar ciertas inferencias, la primera y la principal, la notable brecha que separa la necesidad del mercado laboral sobre la oferta que ofrecen los sectores educativos, en el intervalo de años 2012-2015 la tasa de graduados era mayor a la de ofertas laborales activas en el mercado, lo que quiere decir, que en esa temporada, la dinámica de empleo estaba en un punto más equilibrado que en el futuro, en el punto crítico en la brecha de este intervalo, la diferencia era de hasta 230 mil, esto en el año 2012, y se vio su regulación y lo que se

puede llamar, el punto de máximo equilibrio, a mediados de 2015, momento de la línea de tiempo en el que las dos curvas se cruzan, es decir, se tuvo el mismo número de graduados como de ofertas en el mercado laboral y es a partir de este momento cuando la brecha se inclina totalmente hacia el otro lado, y la brecha que anteriormente mencionamos crecería de forma moderada inversamente. después del año 2015, se empieza a desarrollar una moderada deficiencia general de oferta profesional, como se expuso anteriormente el ritmo de crecimiento anormal que tiene la curva de demanda laboral muestra un muy rápido incremento anual en la cantidad de ofertas que ofrece el mercado, deficiencia que claramente el índice de 7% de crecimiento anual que tiene la curva de graduados, no podría soportar durante mucho tiempo, se tendría que triplicar el número de graduados en el país para poder soportar la demanda que los sectores productivos requieren, ya que, de mantener este ritmo de crecimiento, las necesidades laborales y profesionales del mercado, nunca podrían ser cubiertas y peor aún, la brecha que significa la deficiencia de fuerza laboral del país, crecería de tal forma que en un punto crearía una crisis económica, producto de la falta de fuerza laboral que apoya directamente el desarrollo económico de la región.

### Tendencias según el nivel educativo

El siguiente fragmento que será analizado, es el de la evolución del mercado en términos del nivel educativo que requiera la demanda y ofrezca la oferta. Este análisis muestra indicios que permiten modelar la situación actual del país en cuanto a lo que las universidades están brindando al ciclo de empleo, ¿Qué nivel de educación es el que más demanda el mercado? ¿Qué nivel educativo es el que más gradúa estudiantes en Colombia? son preguntas que se responderán en este fragmento de estudio.

A continuación, se muestran dos gráficas de distribución, que muestran como históricamente se ha dividido generalmente la necesidad del mercado y los egresados de centros educativos del país según su nivel educativo:

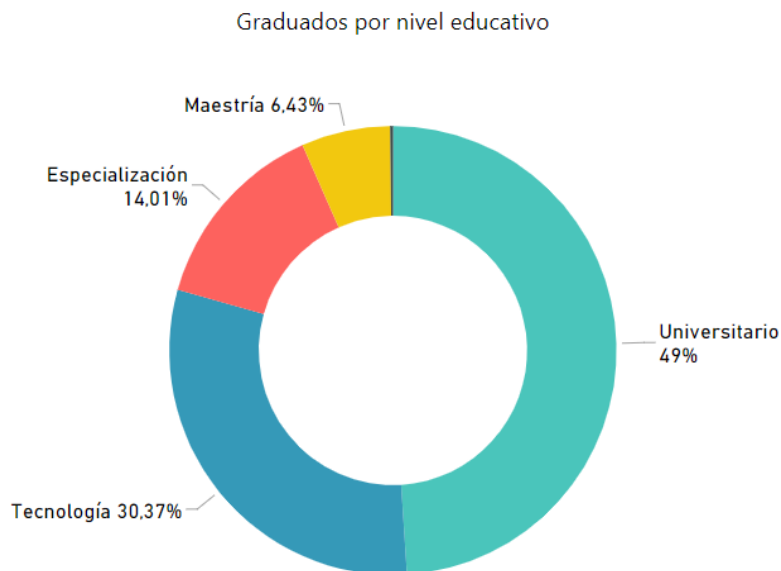
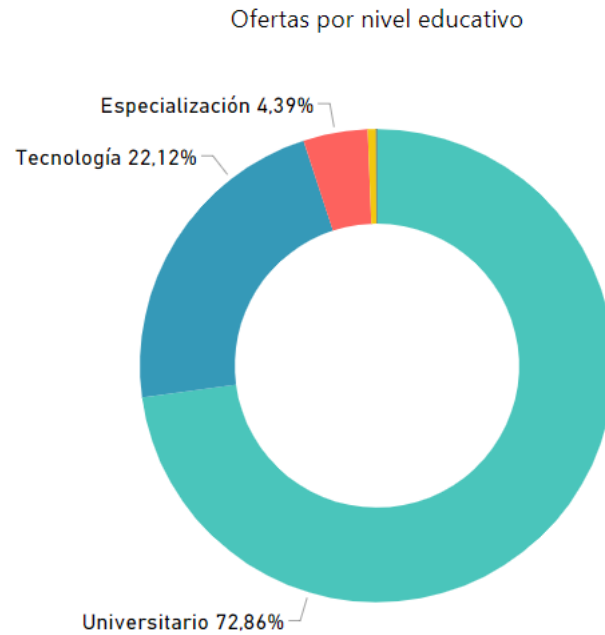


figura 13. Egresados históricos nacionales según nivel educativo



*figura 14.* ofertas históricas por nivel de formación

Como se puede observar en las anteriores gráficas, históricamente en ambas fracciones de dinámica del mercado, la prevalencia son los títulos profesionales o universitarios, aunque en mayor medida las ofertas que rondan el mercado van dirigidas a personas que cuenten con estudios universitarios, las instituciones educativas también ofrecen en su capacidad, la mayor cantidad de profesionales al mercado, seguidos de los títulos tecnológicos y de especialización, que son casi dos tendencias exactas entre ambos factores.

Se observa a continuación cómo se comporta la brecha expuesta en la fig. 11 dependiendo del nivel educativo, en este caso los tres más importantes para el mercado (universitario, tecnológico, especialización):

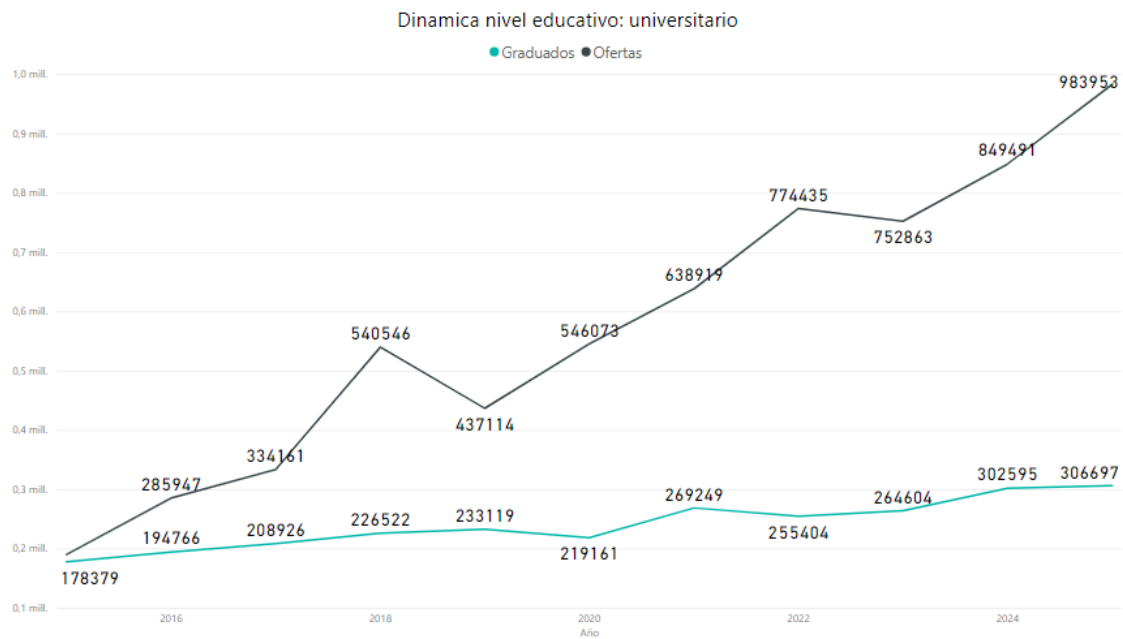


Figura 15. Dinámica de mercado por nivel educativo: universitario

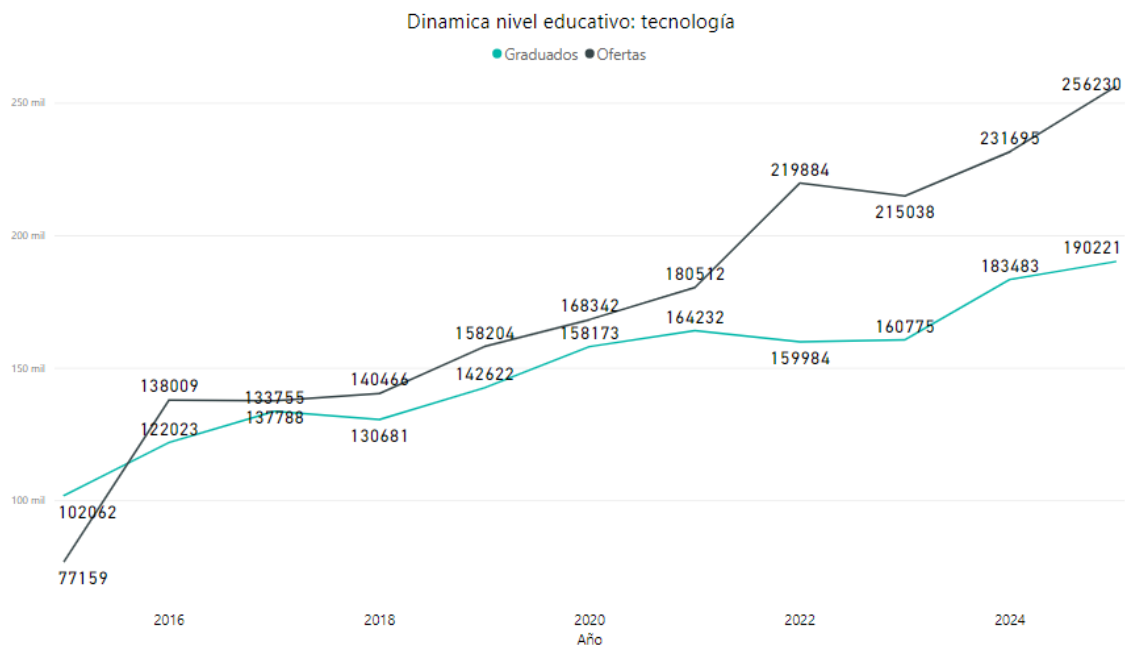


Figura 16. Dinámica de mercado por nivel educativo: Tecnología



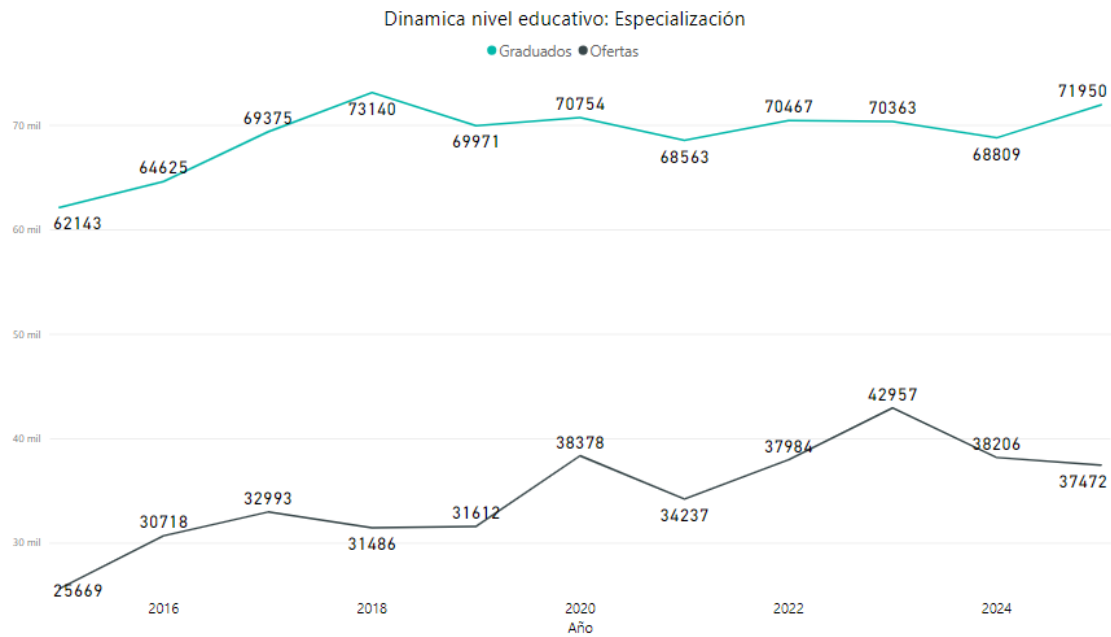


Figura 17. Dinámica de mercado por nivel educativo: Especialización

Estas tres gráficas muestran una ocurrencia interesante, como se puede observar en la fig. 16 (Dinámica de mercado Especialización), ambas curvas tienen un comportamiento casi parecido, tienen en este caso índices de crecimiento muy similares, pero la brecha es demasiado alta, en la mayoría de años, los estudiantes que se gradúan de una especialización, tienen que competir en un mercado que tiene la capacidad de acoger únicamente a la mitad de ellos, aunque este nivel educativo es el único que se satisface en el mercado laboral, como se observa en las fig. 14 y 15, dinámicas de mercado por nivel de formación tecnológica y universitaria, el panorama es distinto, empezando por la dinámica de oferta y demanda de tecnólogos en país, la brecha es leve, hasta el 2016 el mercado se vio exitosamente cubierto, y hasta el 2018 los egresados con este título no tenían una competencia marcada por una deficiencia de empleo, es entonces, a partir del

año 2019 que la oferta laboral cambia de ritmo de crecimiento anual y empieza aumentar a un ritmo mayor a la que lo hace la curva de egresados, la brecha empieza a aumentar a un promedio de 3% anual, lo que deja como resultado la proyección de la deficiencia de tecnólogos para el año 2025, unos 72,000 egresados serían necesarios para cubrir la demanda del mercado en el país para dicho año.

Finalmente tenemos la gráfica mostrada en la fig. 14, la demanda y oferta laboral para estudiantes de nivel universitario o profesional en el país, gráfica que representa una deficiencia grave en la capacidad que tienen los índices de egresados profesionales para cubrir la elevada demanda que ha tenido y tendrá el mercado en el país, para el 2016, año en el que la deficiencia tiene su menor valor, se tiene aproximadamente un 68% más ofertas laborales, que profesionales en el mercado, mientras que la proyección arroja una insuficiencia de fuerza laboral para el año 2025 de hasta un 300%.

En este punto del análisis cabe preguntarse por los índices de desempleo del país que actualmente superan las dos cifras porcentuales, pero, ¿cómo se ve reflejado el desempleo en la medición que acabamos de ver, si lo que logramos plasmar fue una clara deficiencia de oferta laboral? es la primera pregunta que asalta a cualquier persona que se tope con las proyecciones que hemos hecho, la respuesta por su parte, no es muy clara, lo primero que tendríamos que analizar sería entonces, la confiabilidad de los conjuntos de datos utilizados para procesar lo que llamamos “demanda laboral”, como se ha visto, las bases de datos contiene cifras de empleo demasiado altas y parecen no concordar con los datos económicos y desarrollo actuales del país, sin embargo, la información que estas brindan, es extraída de un portal oficial del gobierno nacional, concretamente provienen

del ministerio de trabajo, entonces se ha pensado en otra manera de explicar este comportamiento, y se ha llegado a la teoría de que en general, las instituciones educativas no ofrecen lo que actualmente necesita el mercado, hablando no solamente del número de egresados, sino también de las habilidades y conocimientos que requieren los nuevos paradigmas de empleo que han resultado de la evolución industrial y tecnológica del mundo. Es probable que los profesionales que salen al mercado no cumplan con los perfiles que buscan las grandes compañías o incluso los que necesitan una sección económica en general, conocimientos especializados, habilidades blandas, equipos competitivos y un sin fin de características que se consideran primordiales en el apoyo constructivo de una sociedad.

Entonces, vale la pena plantear, que si las necesidades y los paradigmas en la dinámica laboral del país han evolucionado, también deberían hacerlo las instituciones educativas, es un hecho que los profesionales que están acostumbrados a crear ya no se amoldan al cambio y movimiento que ha experimentado el mundo, las técnicas de enseñanza necesitan ser ajustadas, para integrar en una persona, los conocimientos y habilidades que se necesitan para competir en los principales mercados laborales del país y del mundo

### Áreas de conocimiento y sectores productivos

En esta última parte de análisis se estudió específicamente cada sector económico y área de conocimiento en los que se divide productivamente el mercado laboral y en general la economía del país, las subcategorías de los dos conjuntos principales que describimos al principio, fueron relacionadas y agrupadas paralelamente con el fin de poder realizar comparaciones entre el volumen de graduados que apoya directamente un sector productivo y la demanda que este necesita para funcionar de manera óptima.

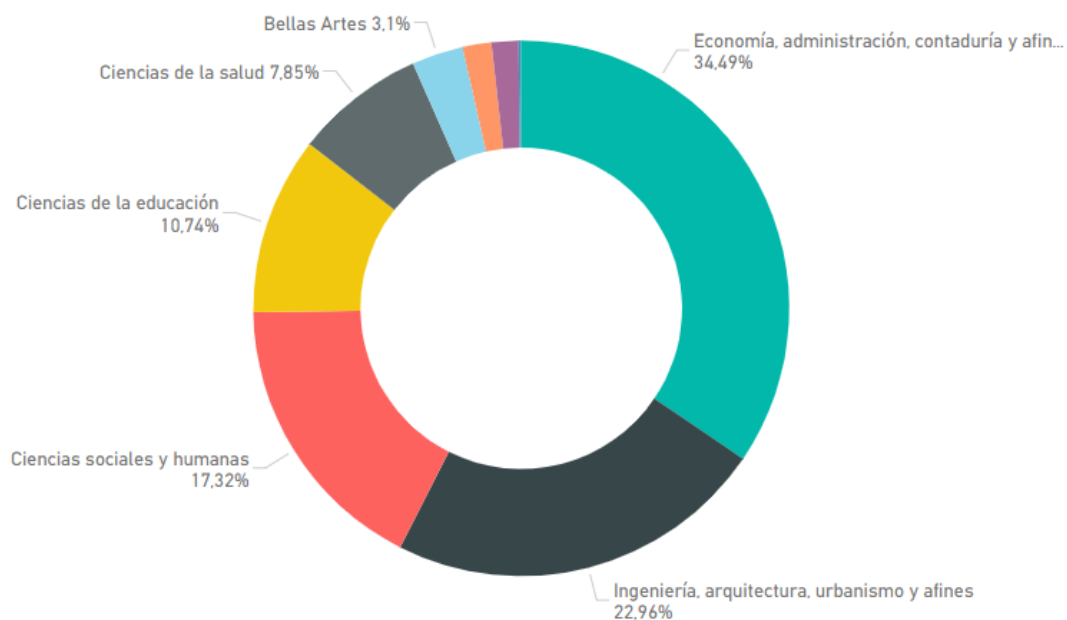


Figura 18. distribución histórica de egresados

## Economía, administración y afines

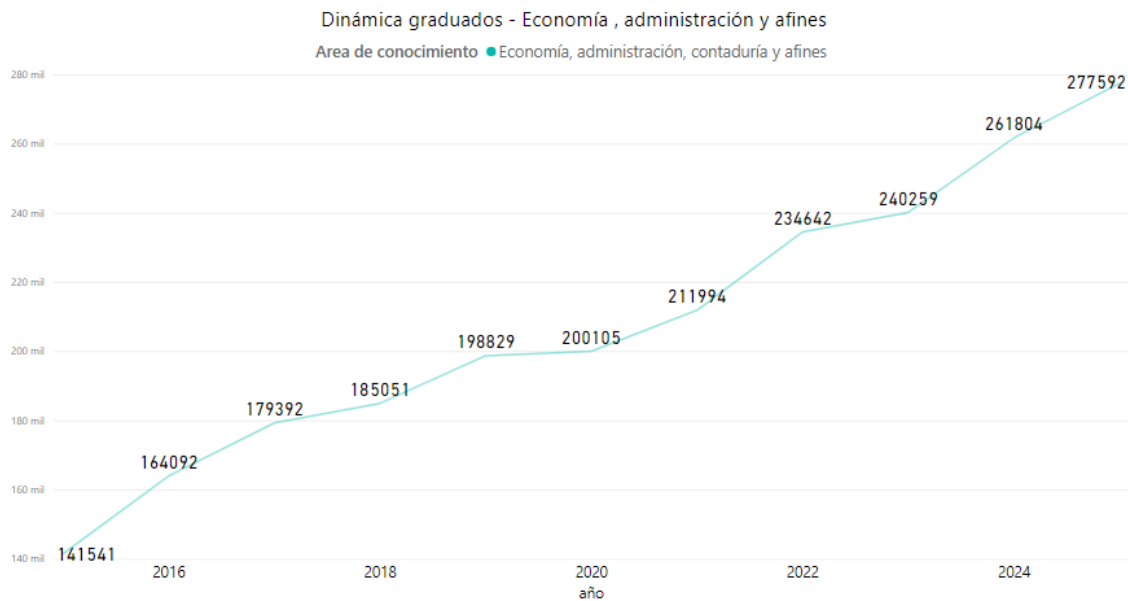


Figura 19. Crecimiento en número de graduados - Economía, administración y afines



Figura 20. Crecimiento en la demanda laboral para el sector economía, administración y afines

La primera sección que se analizará será la que más egresados ofrece al mercado actualmente, la rama de economía y administración, concretamente un 34.49% del total de egresados en el país, son profesionales adscritos a esta rama y como se puede observar en la figura 19, el ritmo de crecimiento anual que esta posee es de 7% anual teniendo un incremento total del 40% con 141.541 estudiantes en el 2015 y 198829 para el 2019. La proyección en esta sección se mantiene estable, para el 2025 se espera tener al menos 277.000 egresados de programas académicos afines con este segmento, cifra, que era de esperarse sería insuficiente para cubrir la demanda de mercado para este sector, ya que la necesidad de la industria por estos tipos de cargos tiene un crecimiento acelerado de 81% anual (véase fig. 20), un ritmo con cadencia mayor a la de egresados, por lo que se encuentra nuevamente una deficiencia moderada de profesionales en el área, actualmente se necesitan cerca de 139.000 economistas o administradores para cubrir satisfactoriamente el déficit que acabamos de describir, adicionalmente se estima que para el 2025 esta brecha aumente hasta los 600.000 egresados.

Este punto se centra en la discusión de cómo se refleja entonces el desempleo en estas cifras, si lo que se evidencia es una situación totalmente contraria, un mercado saturado por oportunidades laborales que un limitado índice de egresados no puede satisfacer.

## Ingeniería, arquitectura, urbanismo y afines

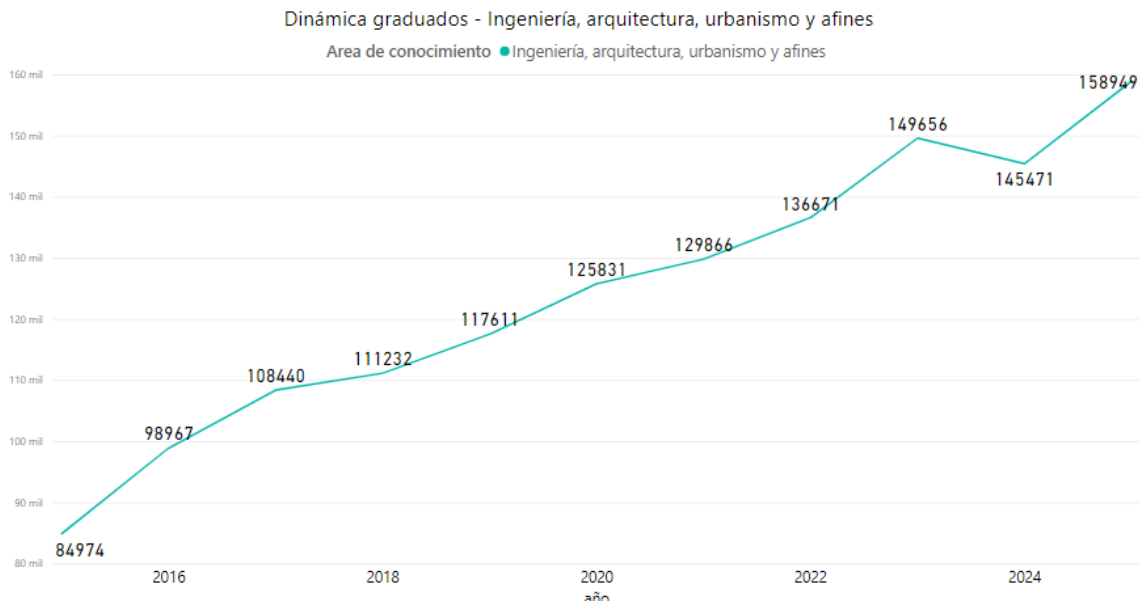


Figura 21. Crecimiento en número de graduados - Ingeniería, arquitectura, urbanismo y afines

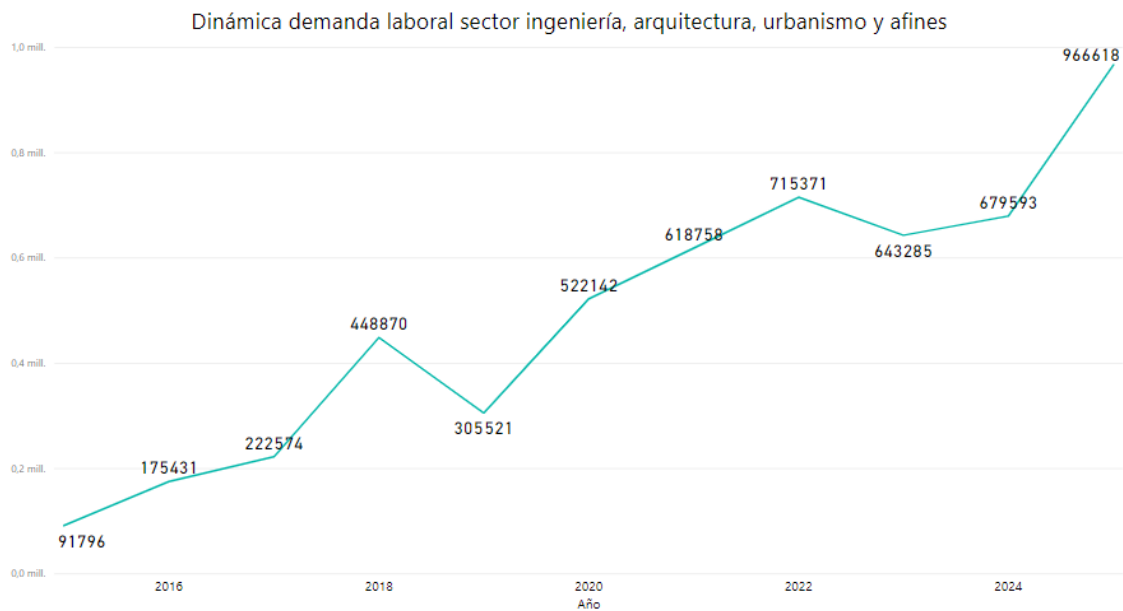


Figura 22. Crecimiento en la demanda laboral para el sector Ingeniería, arquitectura, urbanismo y afines

Para la sección de ingenierías el panorama no es muy distinto, un ritmo de crecimiento en la demanda de 78% genera una brecha de hasta 180.000 profesionales para el 2019 y una estimación de 500.000 para el 2025, ya que el índice de esta área posee un ritmo de crecimiento de apenas 6.5% y ha tenido un aumento general del 38% con respecto al primer año de muestra, 2015, hasta el año horizonte, 2025, un total de 33,000 egresados.

Este número, fuera de ser una tendencia que se ha evidenciado ya en otra rama económica y nivel general, tiene un fundamento plasmado en el comportamiento de mercado actual, y es que, según estudios de dinámica laboral que realizan distintas firmas de estudio del país (Las carreras del futuro, una elección minoritaria en Colombia. (2019, 17 septiembre). Recuperado 22 septiembre, 2019, de <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/cuantos-estudian-ciencia-ingenieria-o-tecnologia-en-colombia-412116> ), la rama de tecnología y comunicaciones, la que es la base de la que llamamos cuarta revolución industrial, es la que un mayor nivel de deficiencia presenta a nivel nacional, el país cuenta con un número limitado de ingenieros capacitados en el área para cubrir una demanda cada vez más incisiva, como ya hemos hablado antes, la tecnificación del mundo ha cambiado muchas dinámicas sociales, la que más interesa en este caso, la transformación que ha tenido que sufrir el mercado laboral para poder dar cumplimiento al desarrollo de nuevos empleos relacionados con tecnología y el reemplazo en el tipo de mano de obra que requieren las grandes industrias para tecnificar sus procesos y hacerlos más eficientes con el fin de



soportar las necesidades de una población en constante cambio. Esta tecnificación supone entonces, el descarte de puestos de trabajo que dejan de ser necesarios o bien pasan a ser muy poco productivos o eficientes en el proceso industrial de cualquier compañía, dando paso a la necesidad constante de personal que les apoye en su propia revolución tecnológica.

Se estudiaron para esta sección los dos principales sectores económicos del país, los que más aportan a la economía y al mercado objeto de estudio, se ha decidido acotar el análisis gracias a que la tendencia que hemos evidenciado en los dos casos anteriores, se repite para los demás sectores y áreas de conocimiento de los cuales se tiene registro, el comportamiento es el mismo, para todos podríamos denotar una brecha, claramente cada una de distinta proporción, entre la demanda y oferta laboral en el país. Más adelante se realizará entonces un análisis general de las posibles causas de este comportamiento en general, empezando por la próxima sección, en la que se estudiará la tendencia de mercado en algunos departamentos del país con el fin de tratar de explicar las razones de la deficiencia que hemos encontrado y su correlación con los elevados índices de desempleo que vivimos actualmente.

## Dinámica general - Departamentos

En esta sección se estudiará el comportamiento de mercado de algunos departamentos del país, con el fin de hallar una explicación a la contradicción entre los resultados de este artículo y la información plasmada en las cifras oficiales del país, ¿por qué por un lado tenemos una deficiencia generalizada de fuerza laboral y por otra un elevado índice de desempleo? Véase a continuación las tendencias de los departamentos que podrían brindarnos una explicación de este fenómeno, en este caso, los 5 departamentos más poblados en Colombia.

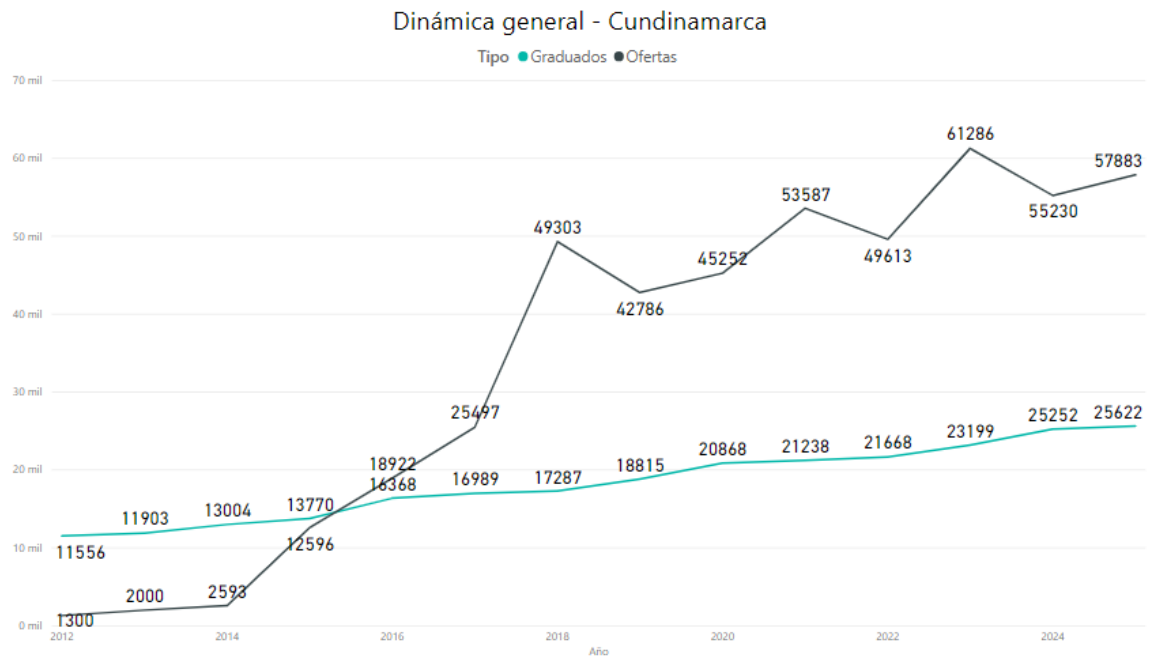


Figura 23. Dinámica de mercado general - Departamento de Cundinamarca

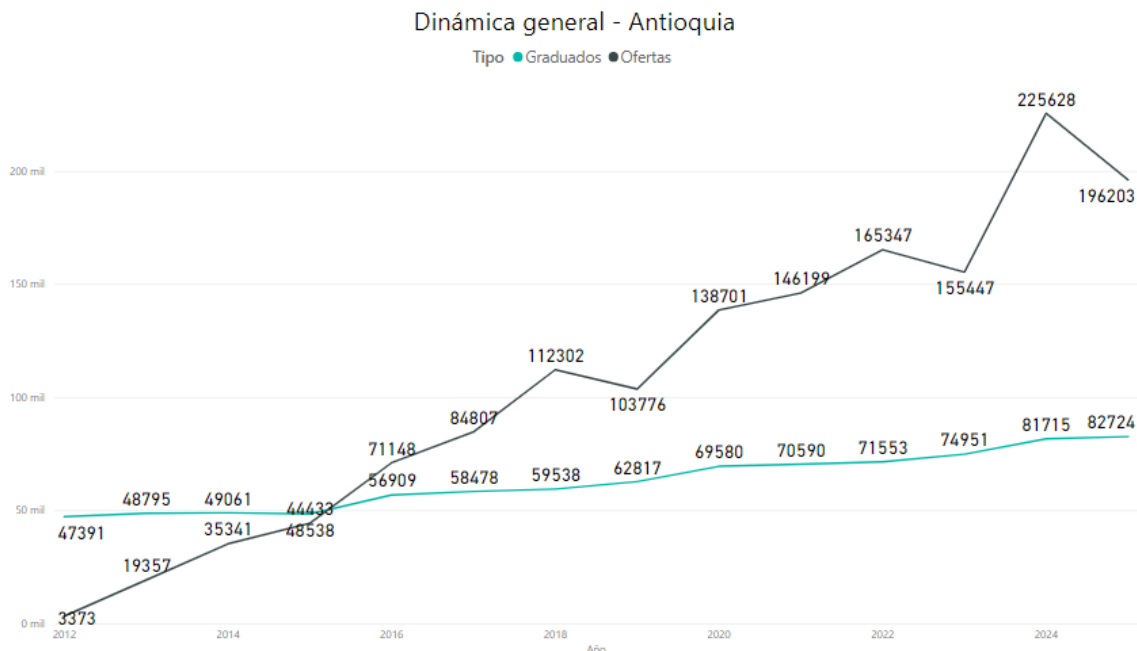


Figura 24. Dinámica de mercado general - Departamento de Antioquia

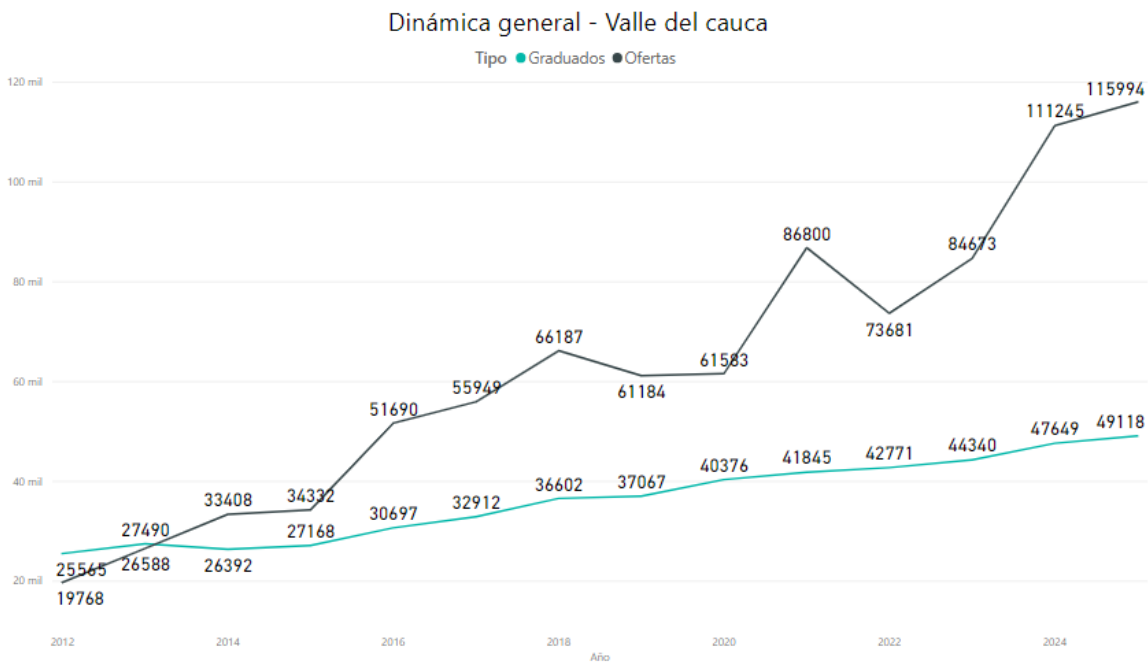


Figura 25. Dinámica de mercado general - Departamento de Valle del cauca

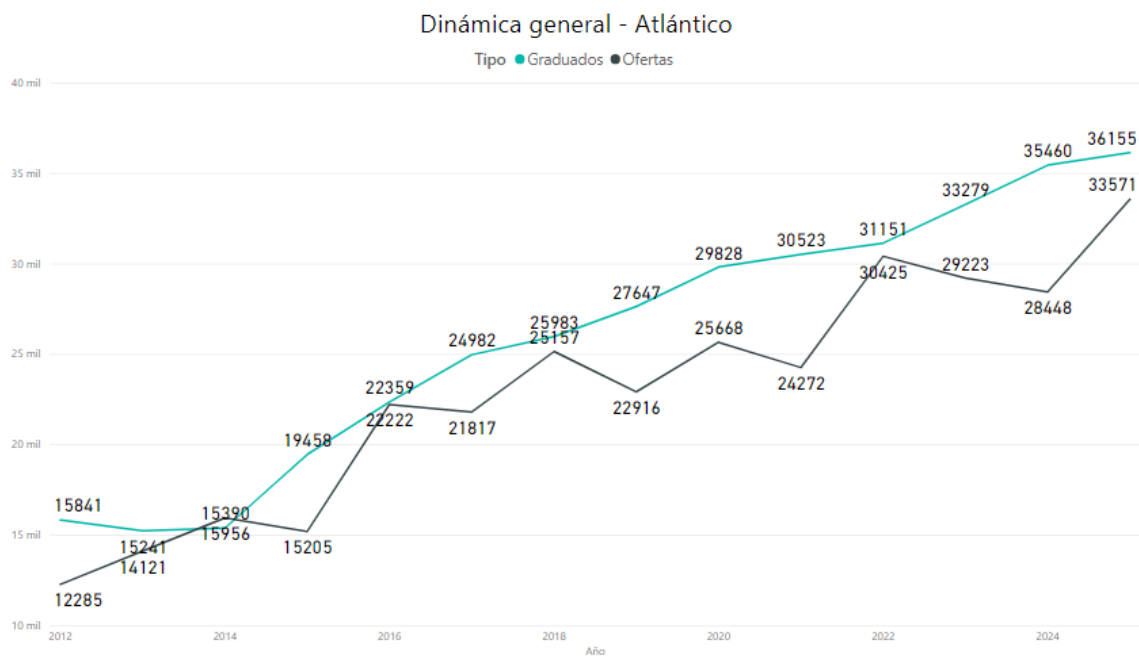


Figura 26. Dinámica de mercado general - Departamento de Atlántico

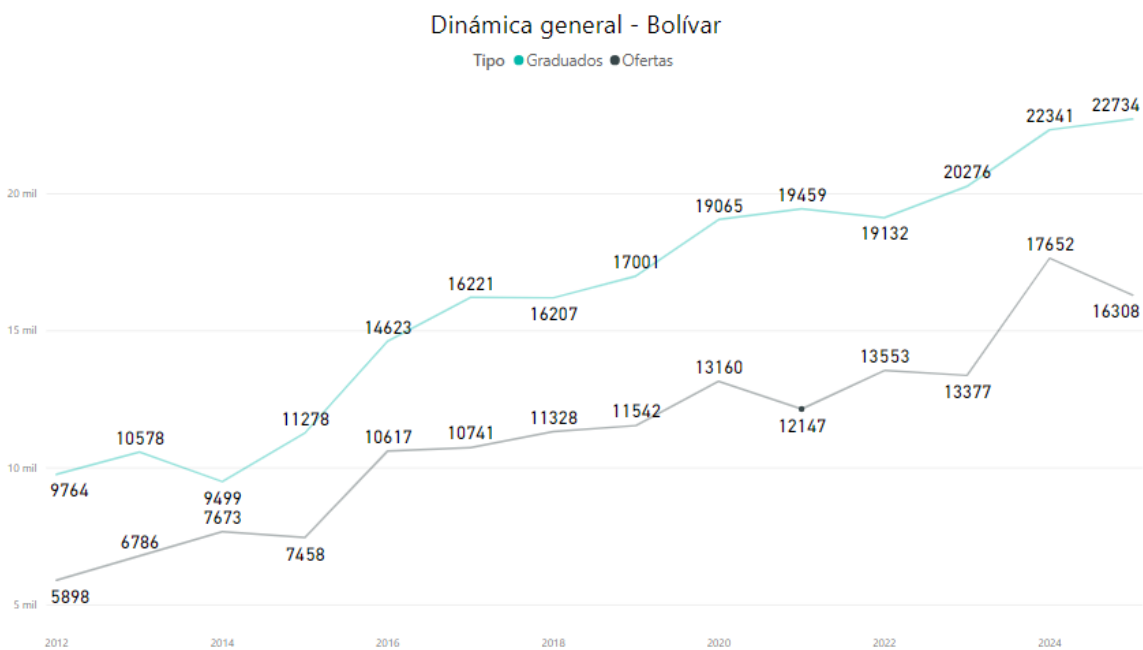


Figura 27. Dinámica de mercado general - Departamento de Bolívar

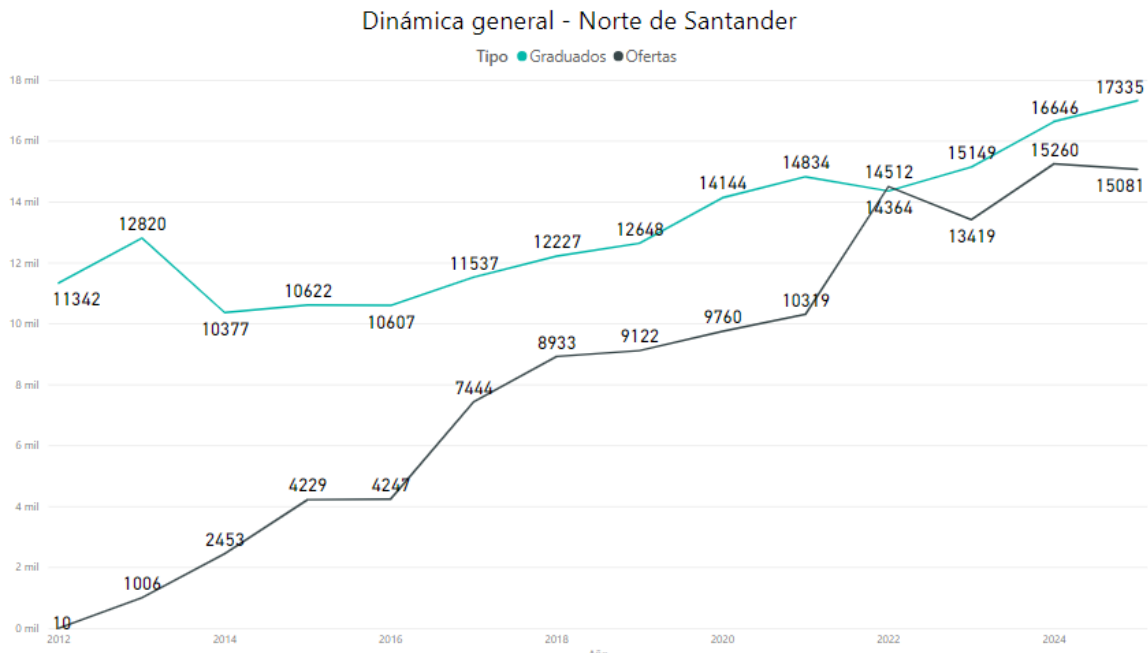


Figura 28. Dinámica de mercado general - Departamento de Norte de Santander

Como se puede observar, en las figuras 23 y 24, correspondientes a la dinámica de mercado general en los dos principales departamentos del país, Cundinamarca y Antioquia, la tendencia es la misma que hemos evidenciado a lo largo del estudio, un número mucho mayor de ofertas laborales que de graduados, por lo que se pudo deducir nuevamente que estos dos departamentos necesitan de muchos más trabajadores para satisfacer las necesidades de su mercado, sin embargo, las tasas de desempleo para ambos se ubicaron en 10,5% y 13,5% respectivamente, entonces, ¿qué es lo que ocasiona este complicado comportamiento? la primera y tal vez más sencilla explicación, es que definitivamente los profesionales del país no cumplen con los perfiles que necesita el mercado, tal vez las habilidades que se aprenden a lo largo de un programa académico actual, no se acoplan a los ajustes que tuvo que sufrir la industria en pro de tecnificar y

mejorar sus procesos para hacerlas más eficientes y cubrir la creciente demanda de bienes y servicios que genera el ritmo de expansión de las comunidades y la población del país y el mundo. Esta premisa estructura la cuestión de si son las universidades los actores que influyen principalmente en el acoplamiento que tienen los profesionales con el mercado laboral al cual tienen salida.

Finalmente, se tienen los índices de los siguientes tres departamentos más poblados, las figuras 26, 27, y 28 muestran el comportamiento de mercado de los departamentos de Atlántico, Bolívar y Norte de Santander, en los cuales se pueden notar, cifras más acordes con los datos oficiales que tenemos por parte del gobierno, para el caso de los 3 departamentos se puede evidenciar una brecha que divide la oferta y la demanda para cada uno, pero esta vez la deficiencia cambia de lugar mostrándonos 3 regiones en las que el número de graduados que interactúan con la demanda laboral es mucho mayor a las ofertas laborales que pueden soportarlos. Pero en este punto ocurre algo interesante, los índices de desempleo en el departamento del Atlántico son de apenas 7.6%, para Bolívar un 5,8% y para Norte de Santander 13,4%. Estos índices muestran un comportamiento atípico, atípico en el sentido lógico, ¿por qué en una ciudad como Bogotá, en la cual hay una mayor concentración de ofertas laborales que de graduados, el desempleo es mayor a otra que tiene en caso contrario, más graduados en su mercado que ofertas? no se puede dar una respuesta concreta, se puede decir que la dinámica social influye en el desarrollo de una región, o simplemente factores más profundos de índole cultural los que provoquen comportamientos de mercado como los que se acaban de ver.

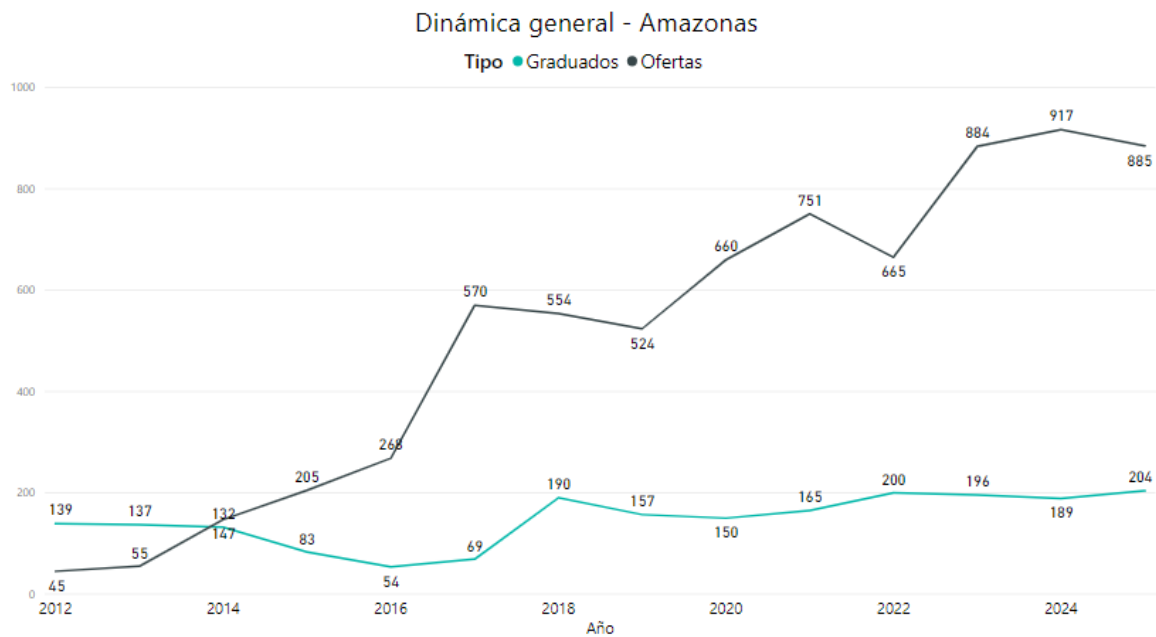


Figura 29. Dinámica de mercado general - Departamento de Amazonas

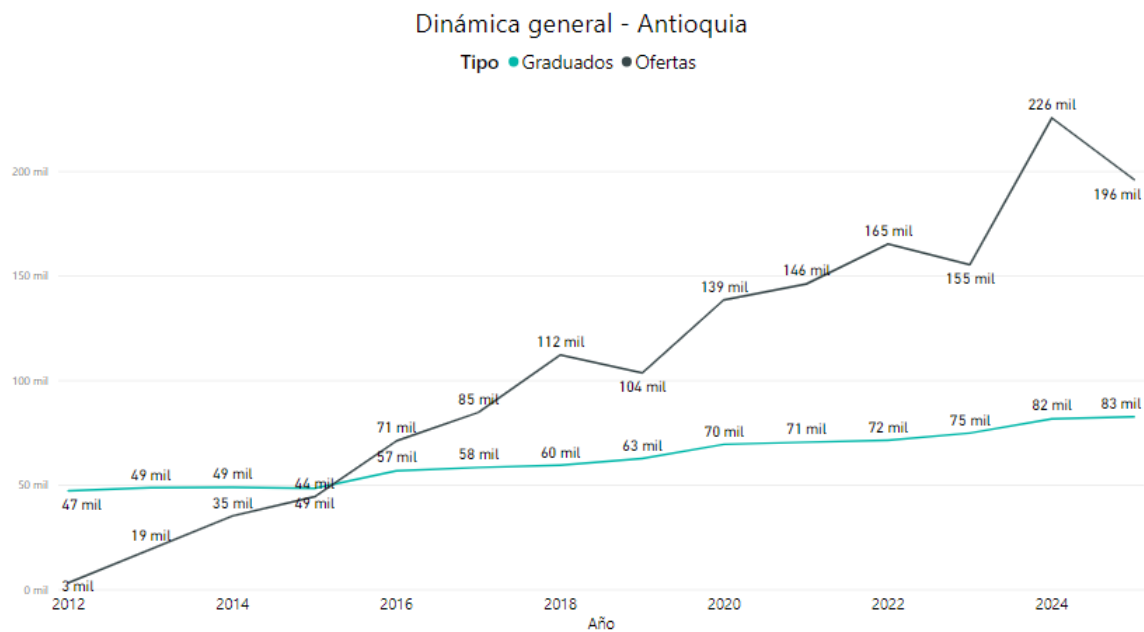


Figura 30. Dinámica de mercado general - Departamento de Antioquia

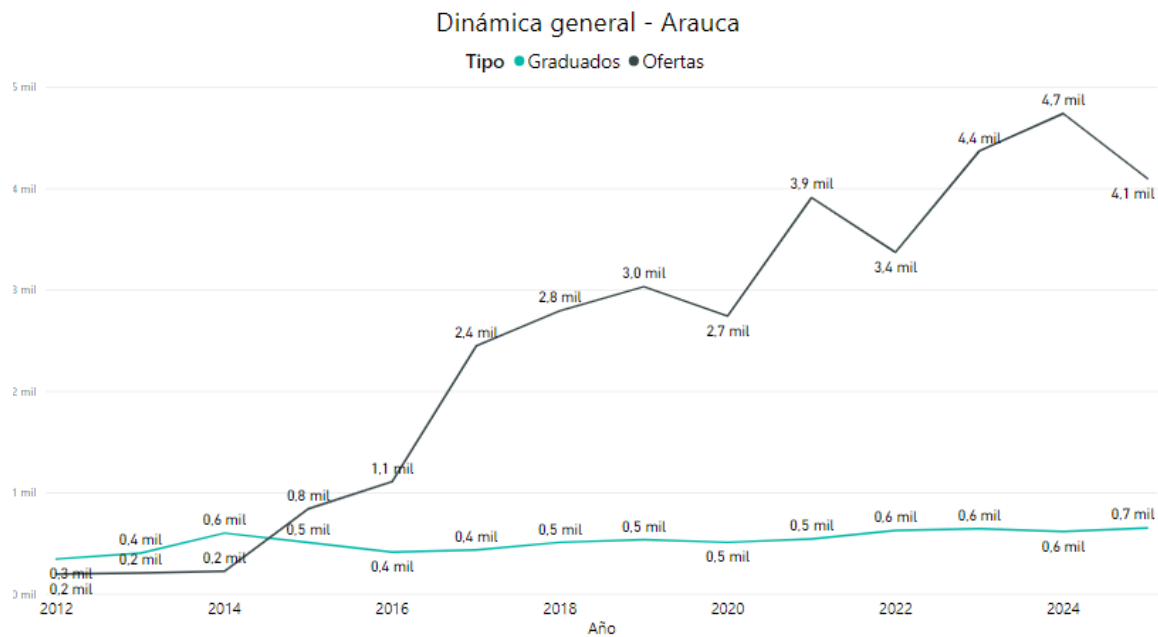


Figura 31. Dinámica de mercado general - Departamento de Arauca

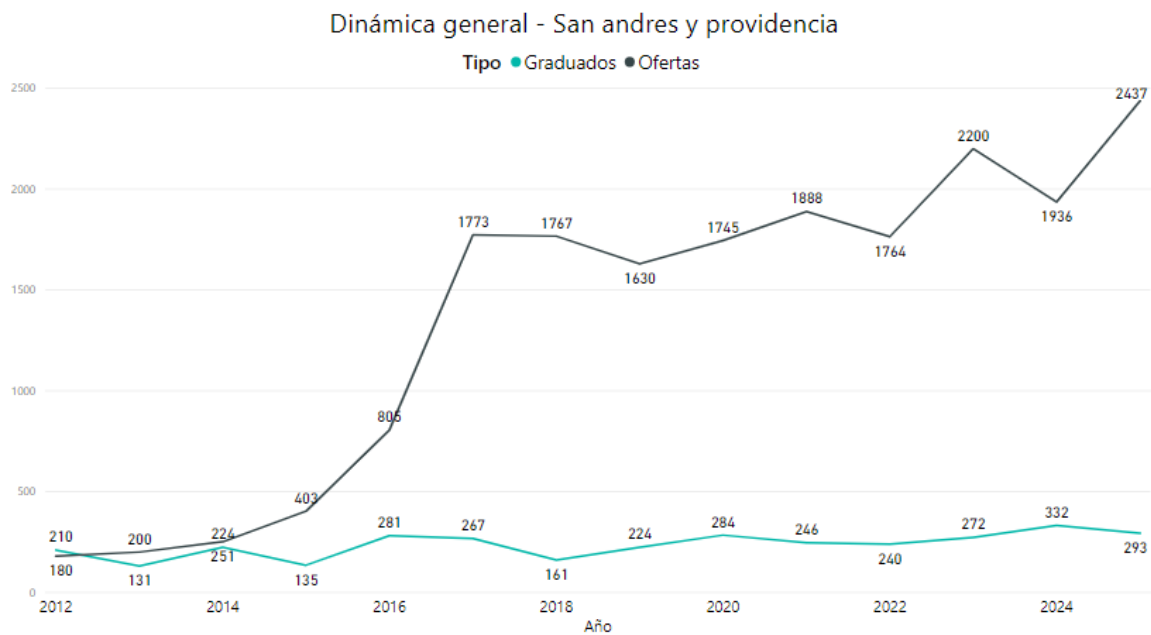


Figura 32. Dinámica de mercado general - Departamento de San Andrés y providencia



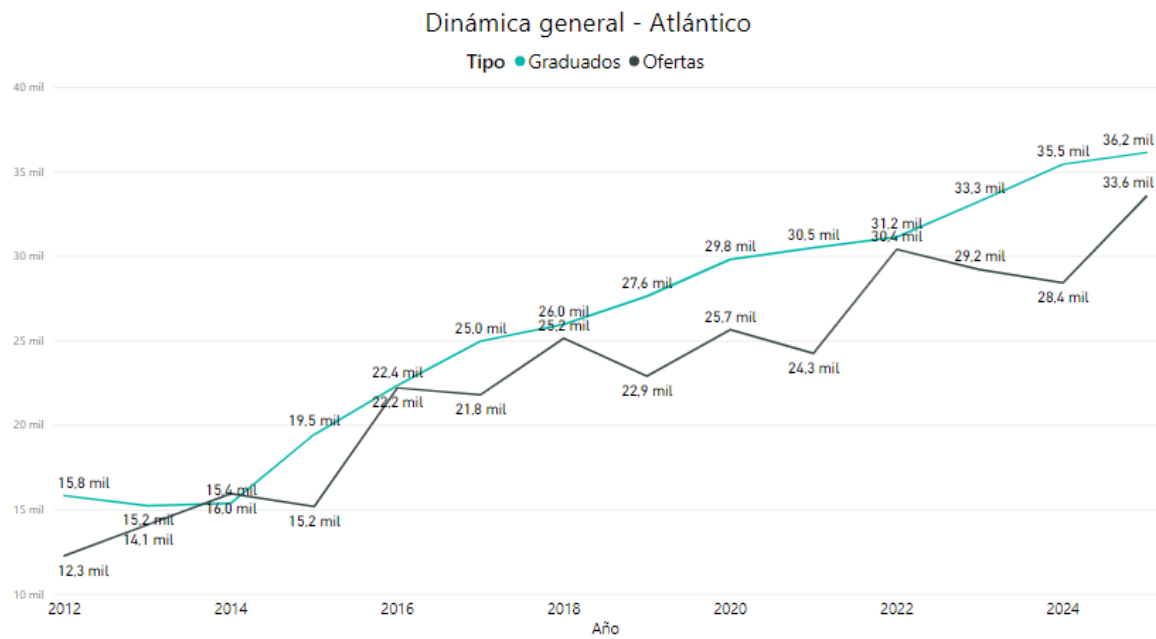


Figura 33. Dinámica de mercado general - Departamento de Atlántico

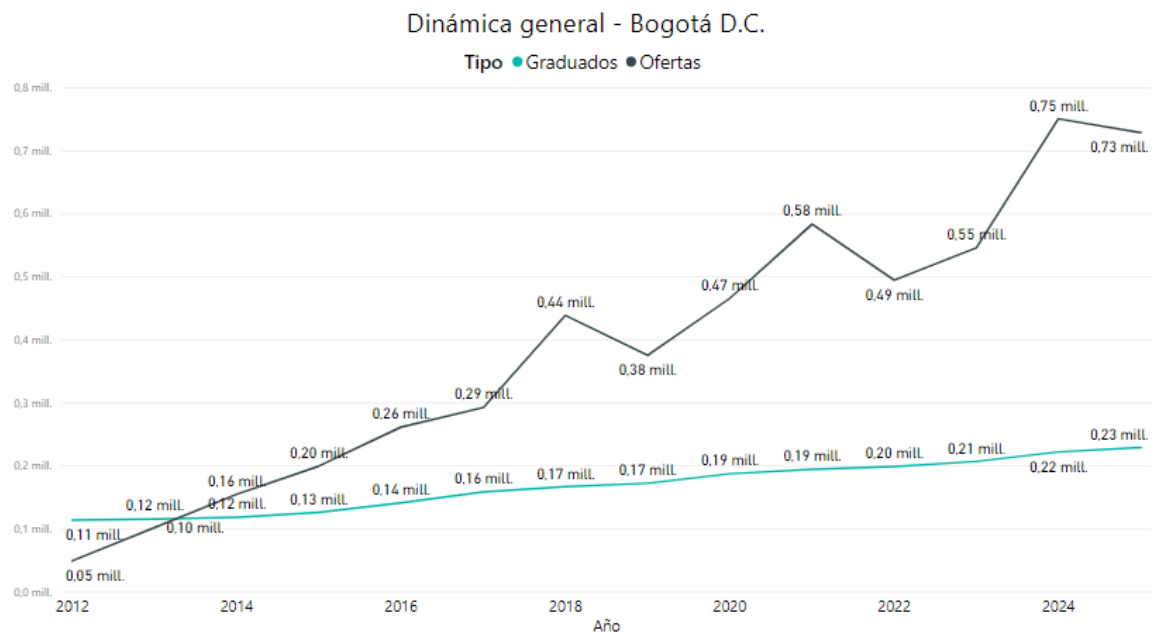


Figura 34. Dinámica de mercado general – Bogotá D.C.

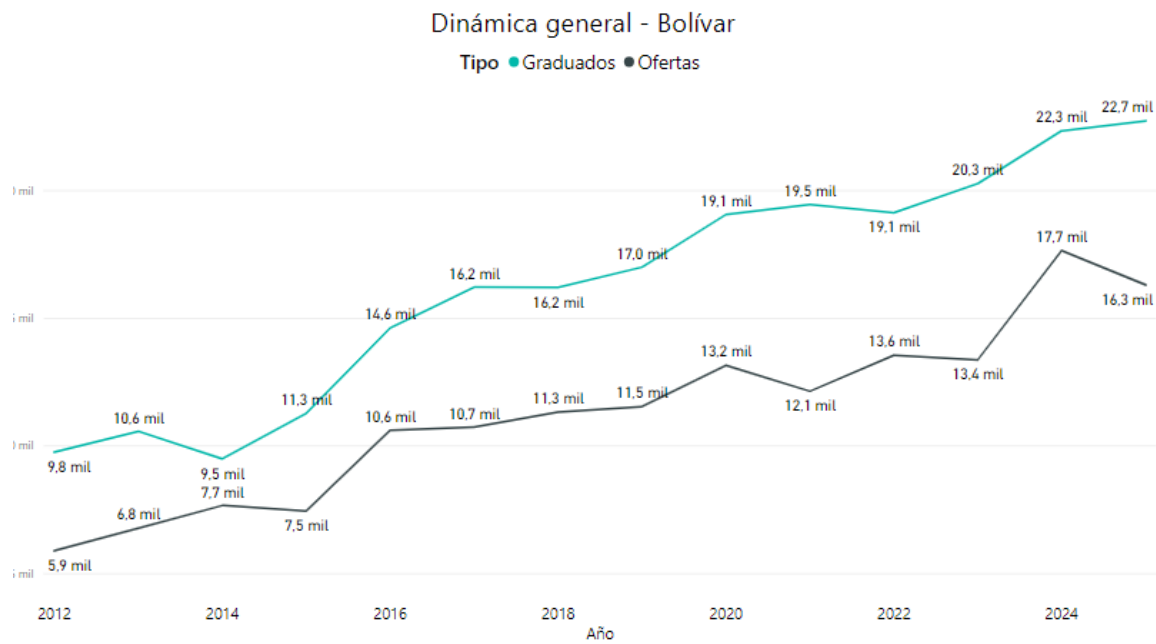


Figura 35. Dinámica de mercado general - Departamento de Bolívar

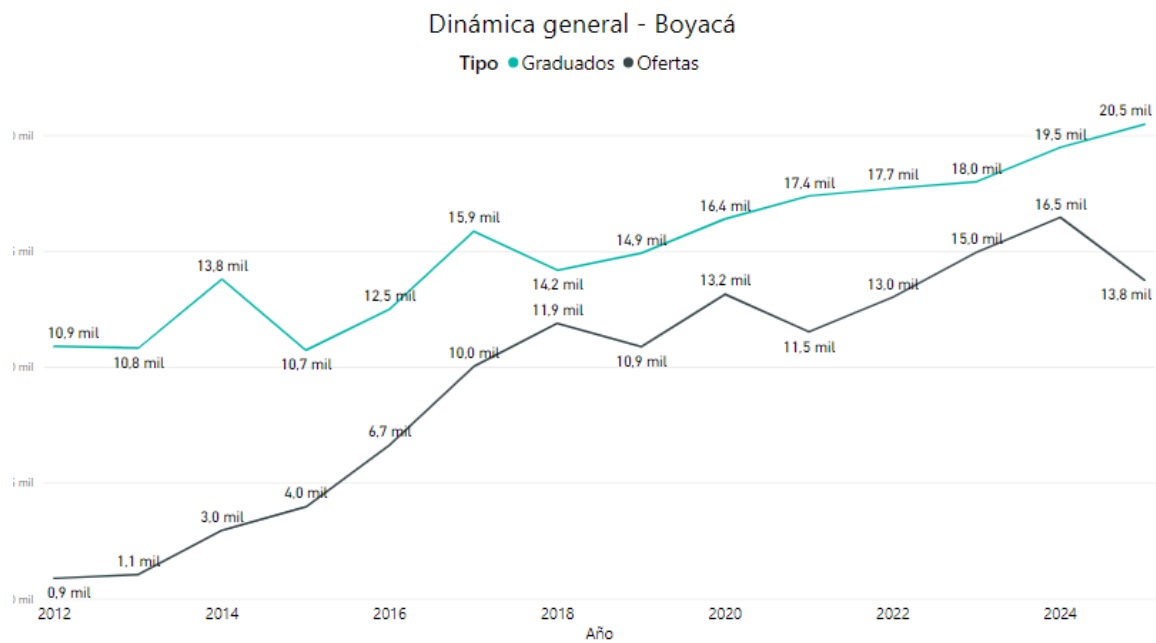


Figura 36. Dinámica de mercado general - Departamento de Boyacá

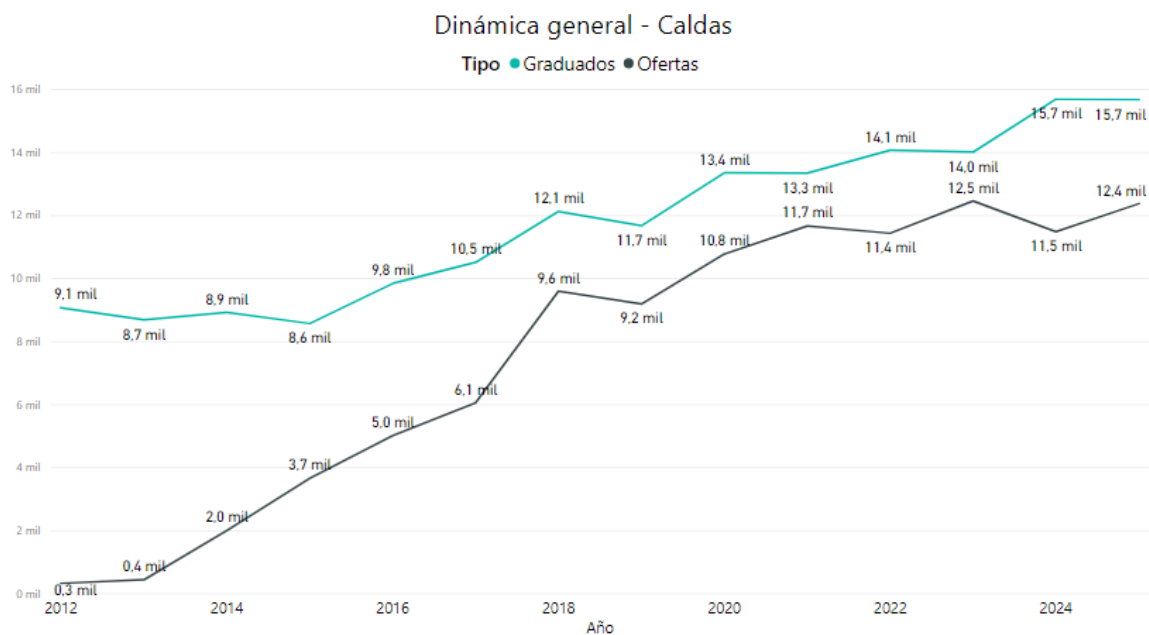


Figura 37. Dinámica de mercado general - Departamento de Caldas

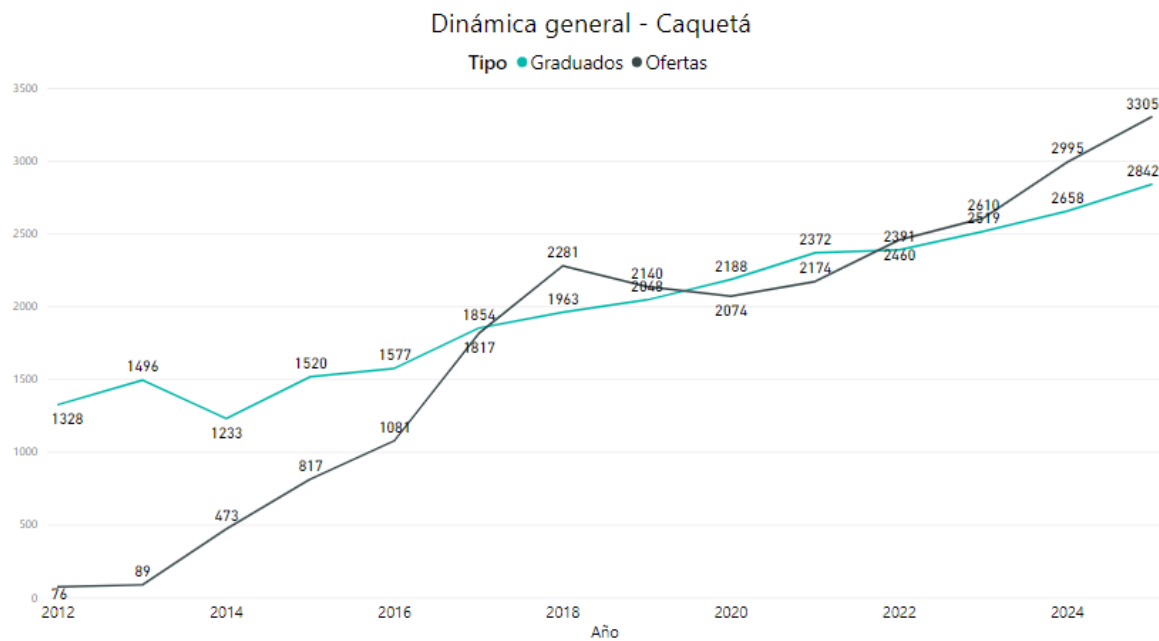


Figura 38. Dinámica de mercado general - Departamento de Caquetá

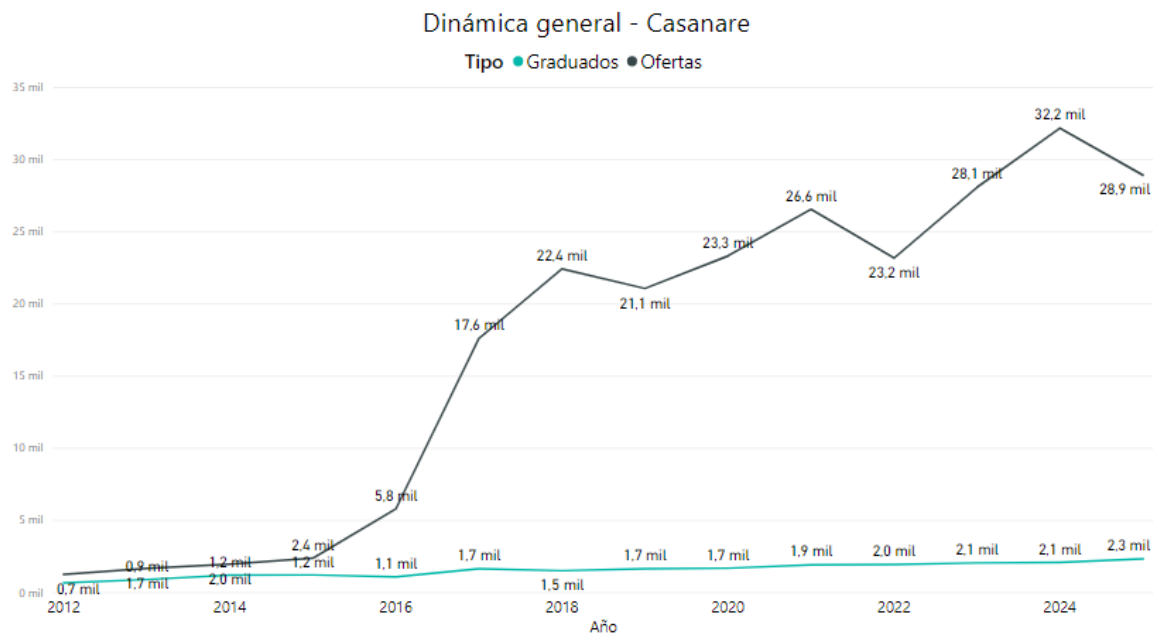


Figura 39. Dinámica de mercado general - Departamento de Casanare

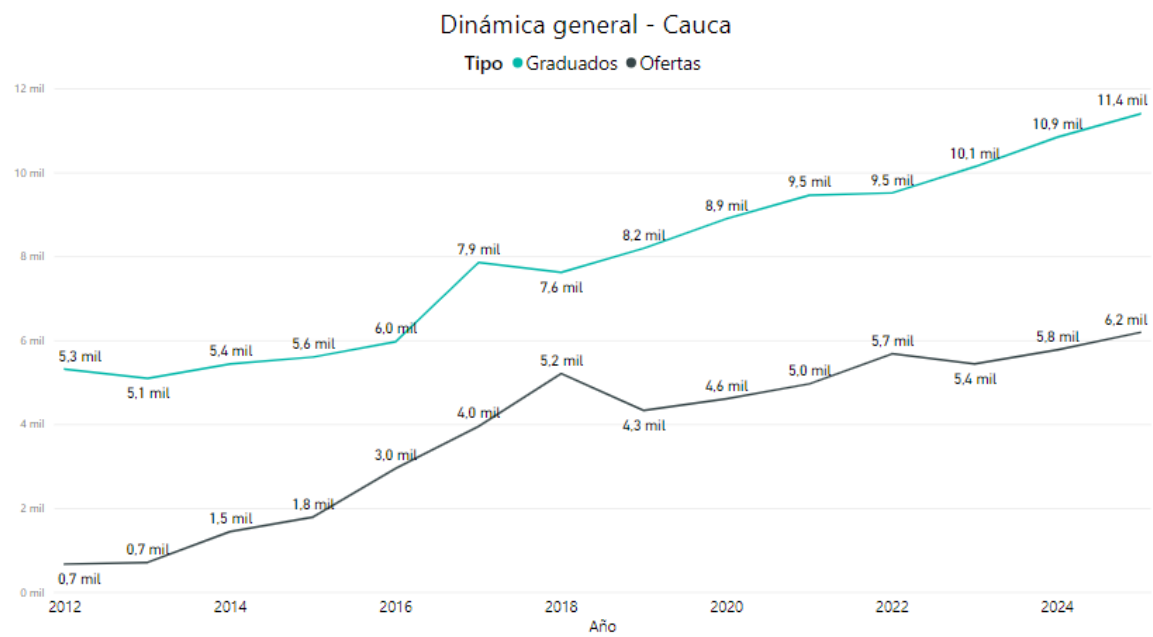


Figura 40. Dinámica de mercado general - Departamento de Cauca

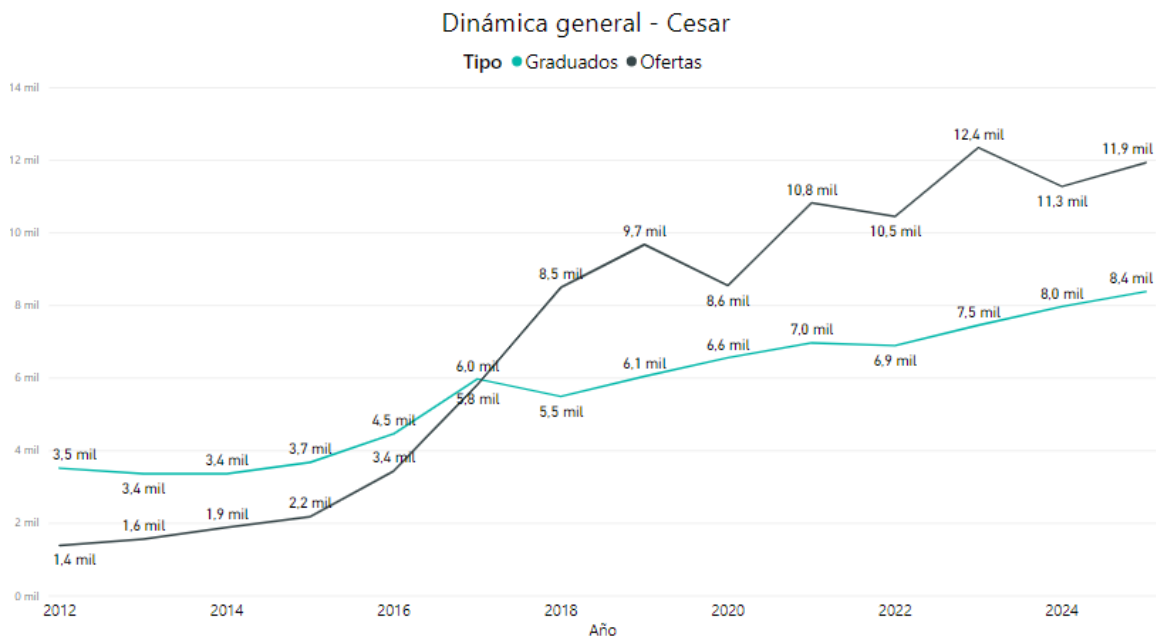


Figura 41. Dinámica de mercado general - Departamento de Cesar

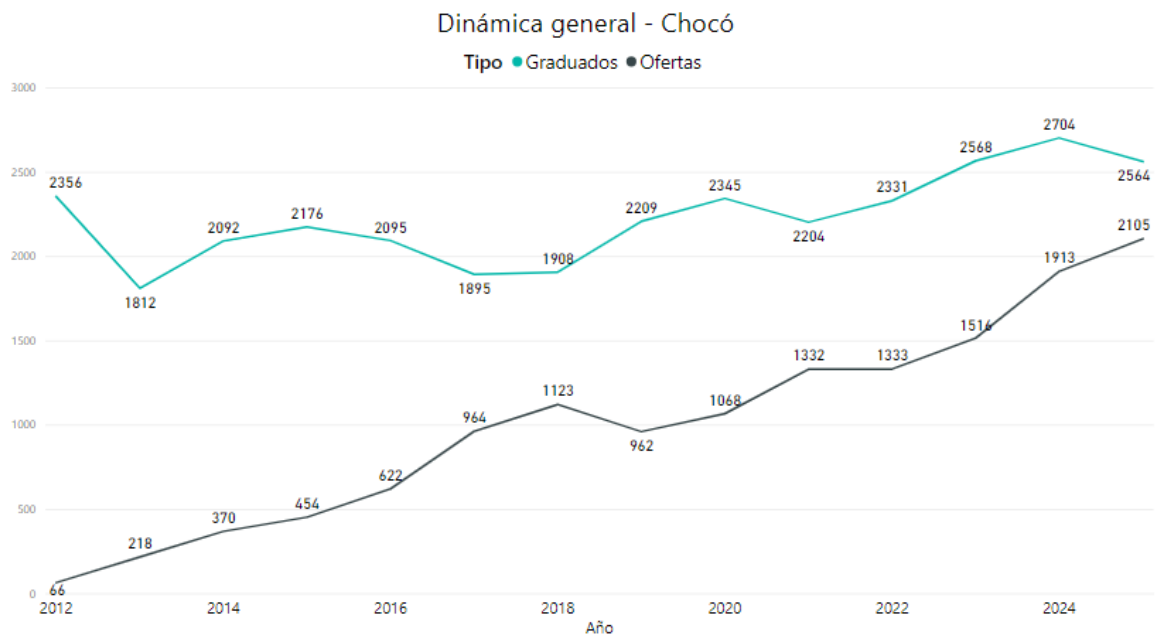


Figura 42. Dinámica de mercado general - Departamento de Chocó

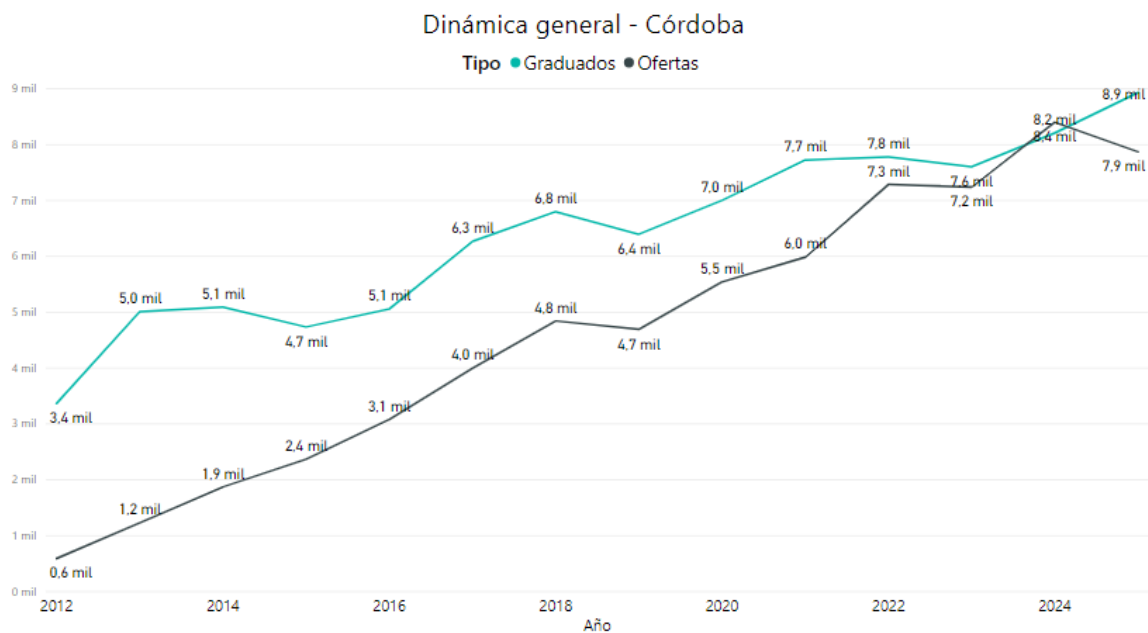


Figura 43. Dinámica de mercado general - Departamento de Córdoba

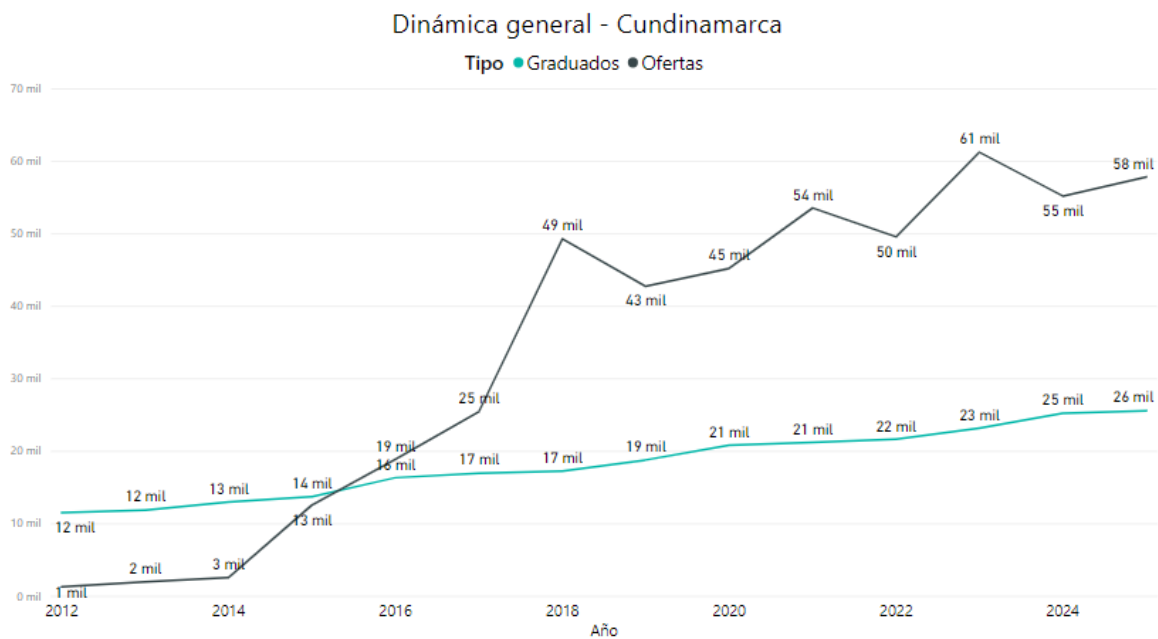


Figura 44. Dinámica de mercado general - Departamento de Cundinamarca

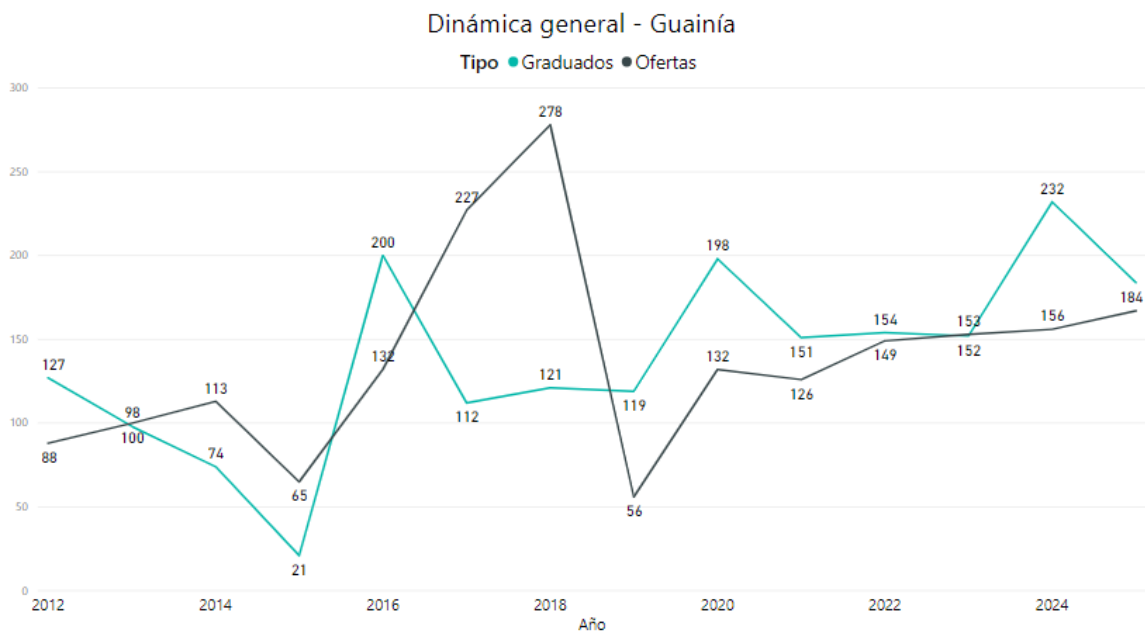


Figura 45. Dinámica de mercado general - Departamento de Guainía

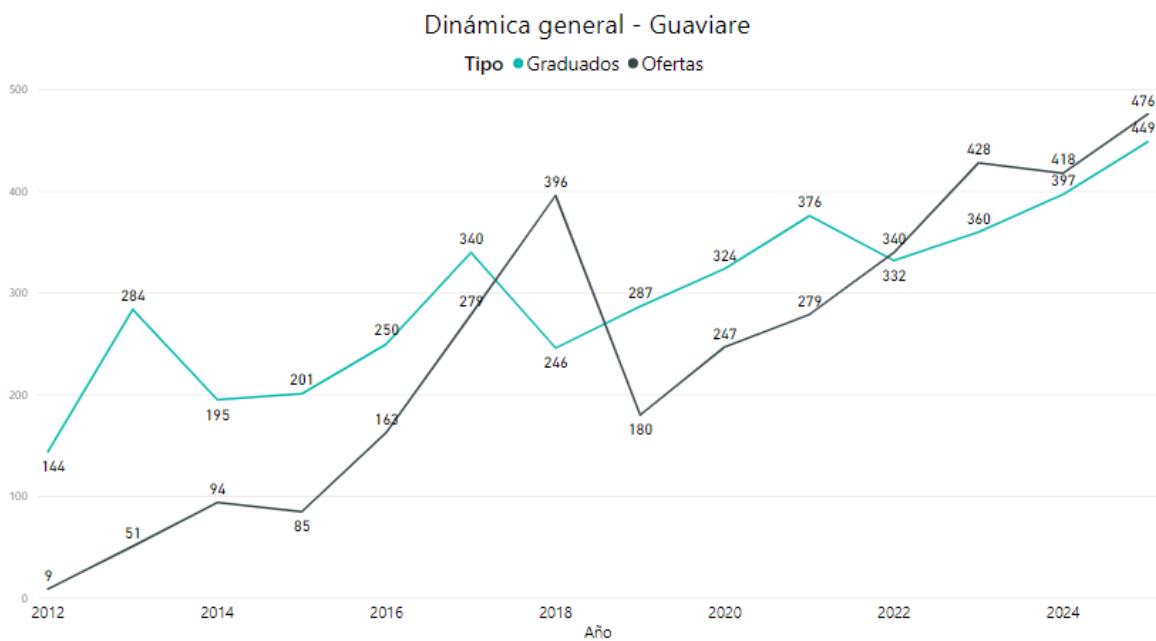


Figura 46. Dinámica de mercado general - Departamento de Guaviare

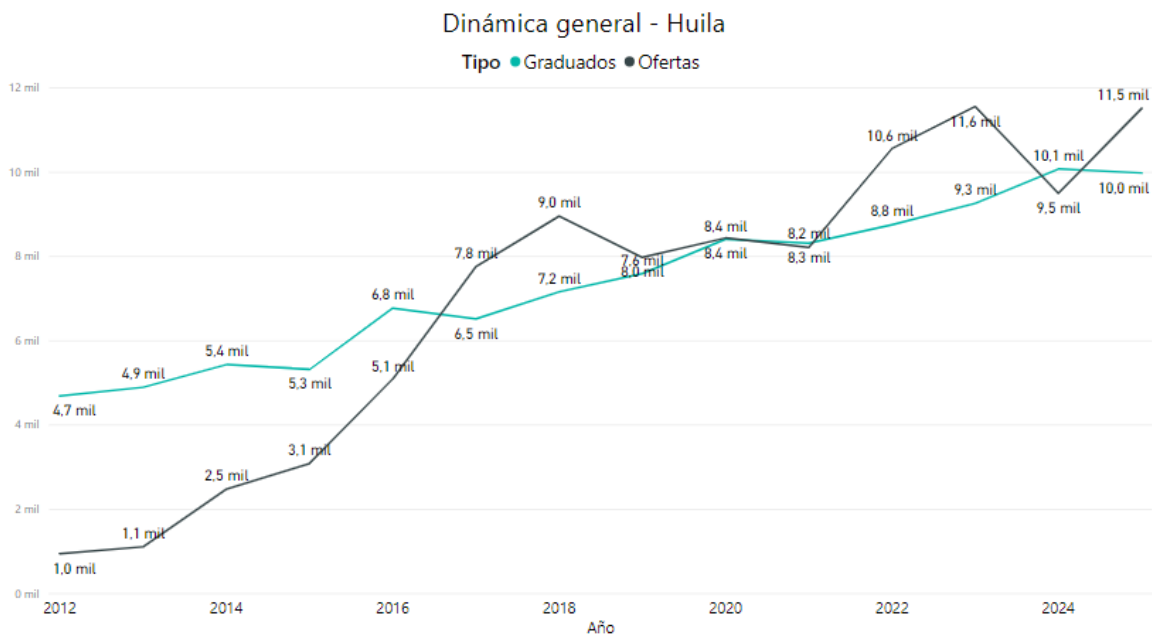


Figura 47. Dinámica de mercado general - Departamento de Huila

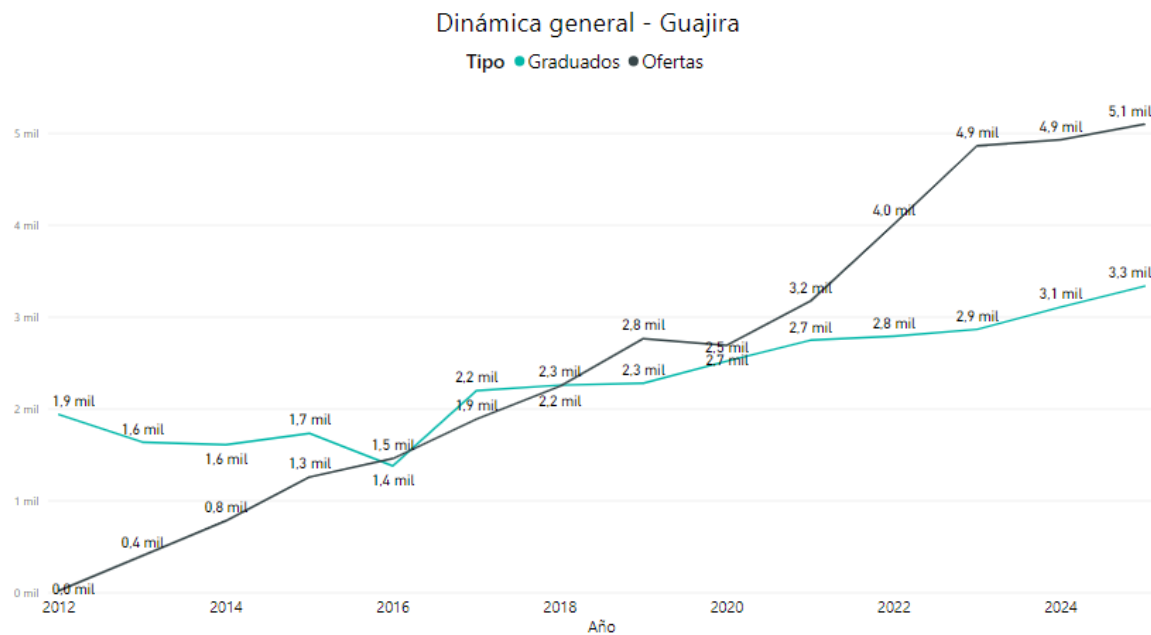


Figura 48. Dinámica de mercado general - Departamento de Guajira



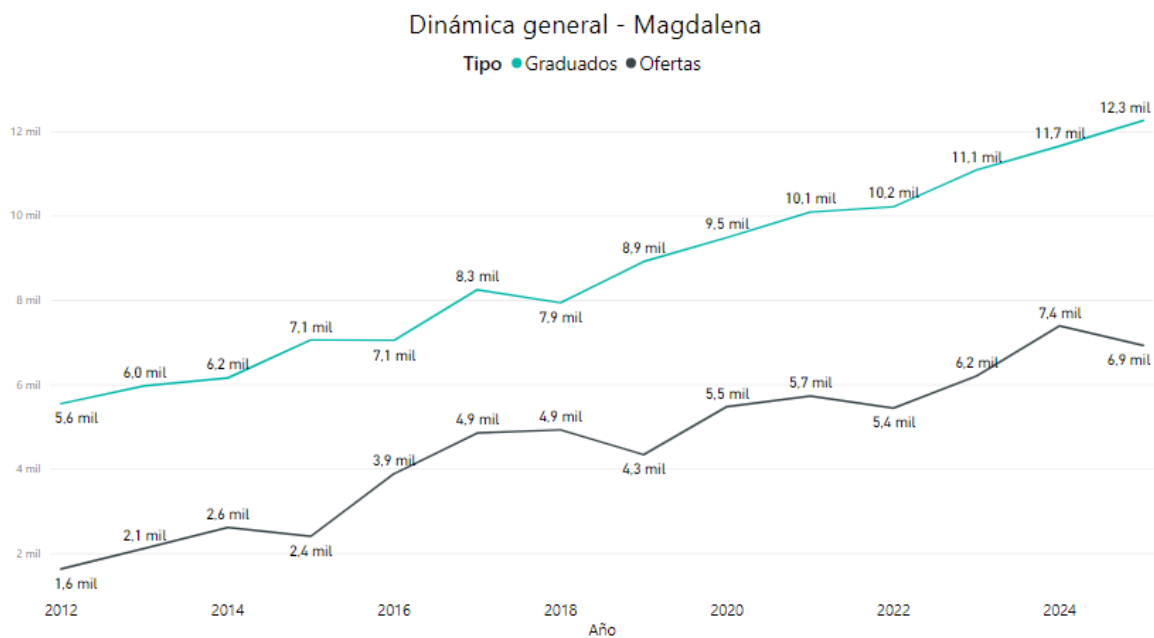


Figura 49. Dinámica de mercado general - Departamento de Magdalena

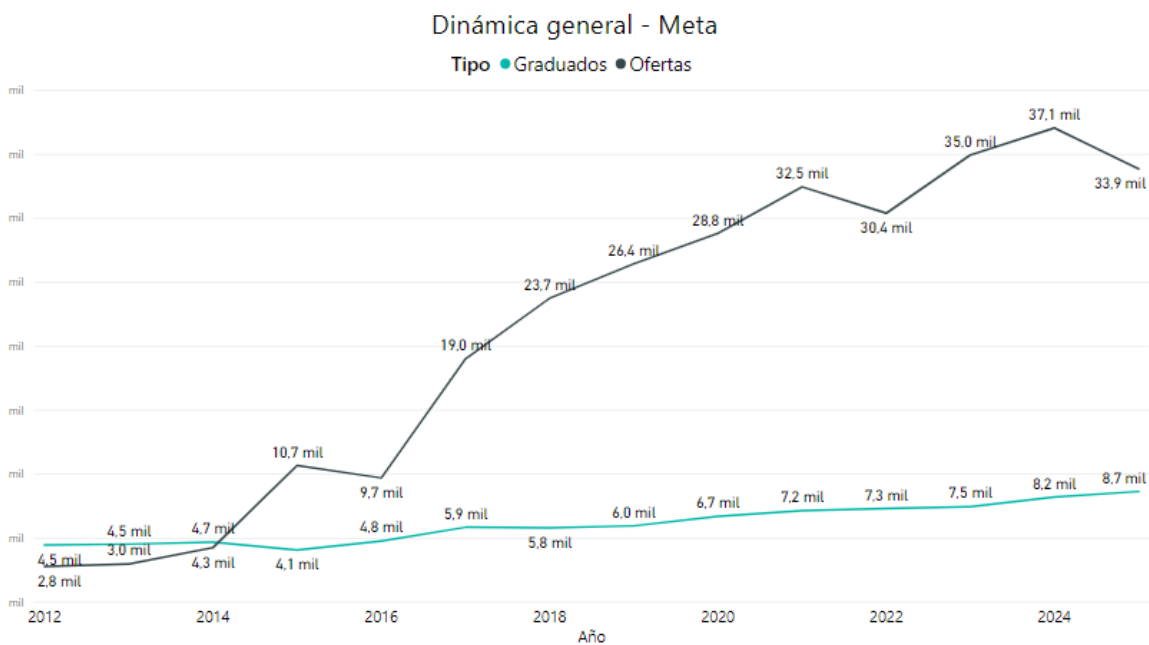


Figura 50. Dinámica de mercado general - Departamento de Meta

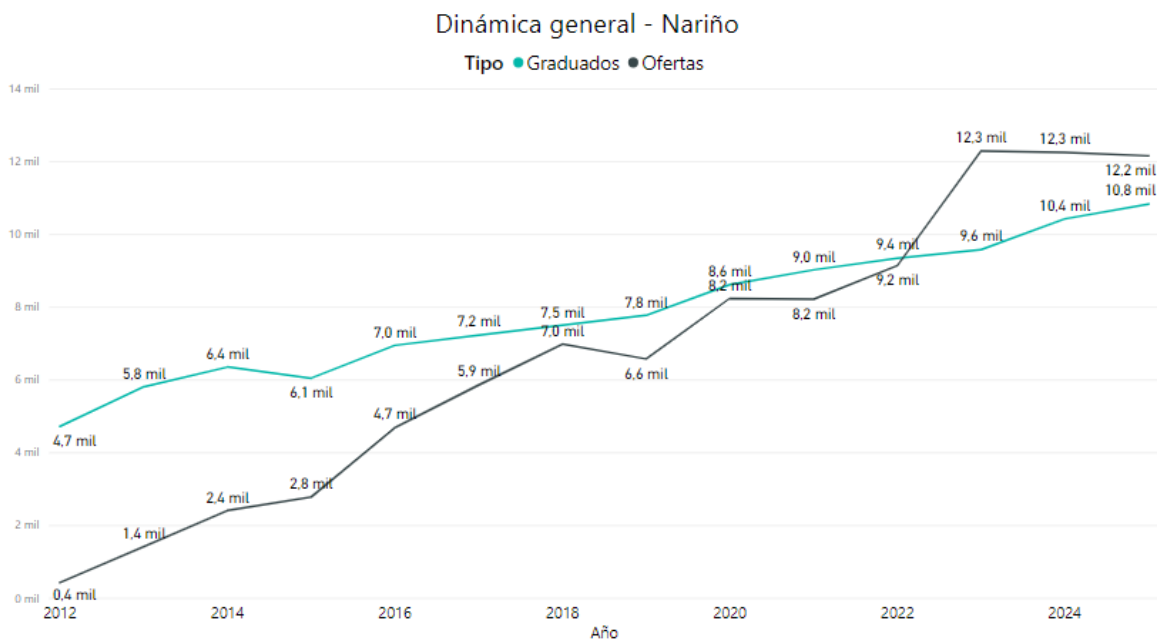


Figura 51. Dinámica de mercado general - Departamento de Nariño

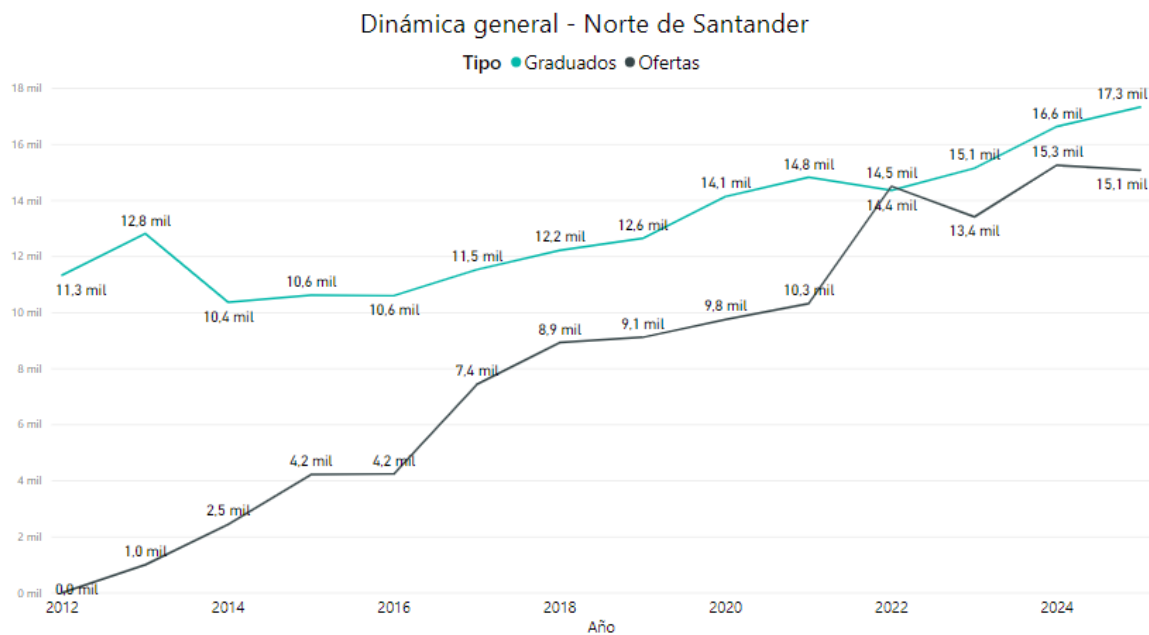


Figura 52. Dinámica de mercado general - Departamento de Norte de Santander

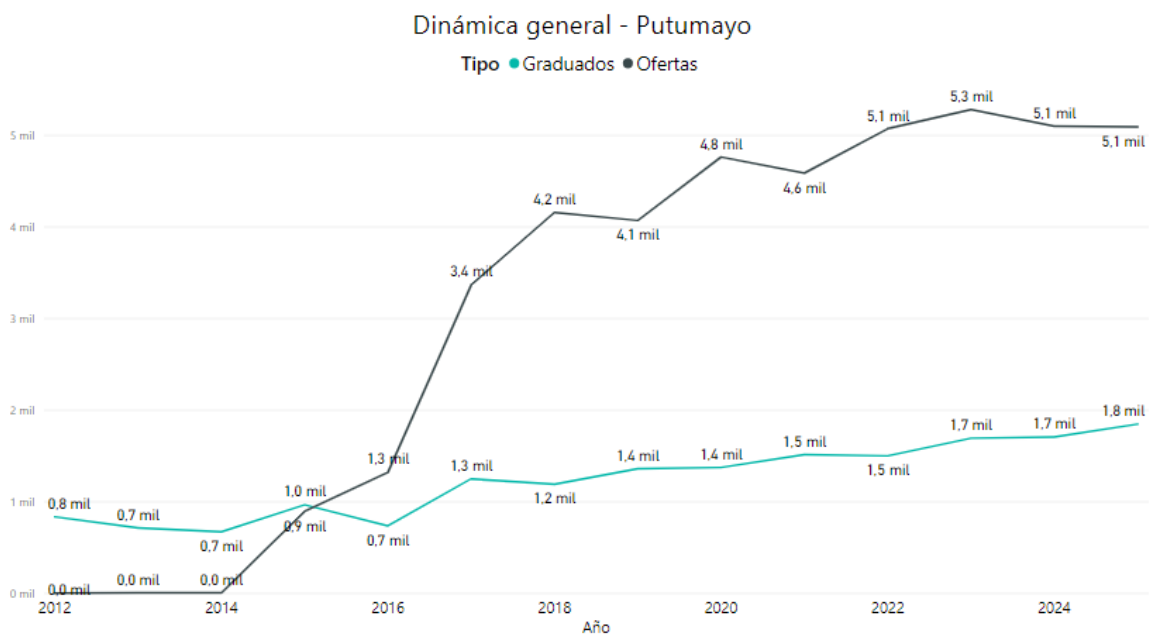


Figura 53. Dinámica de mercado general - Departamento de Putumayo

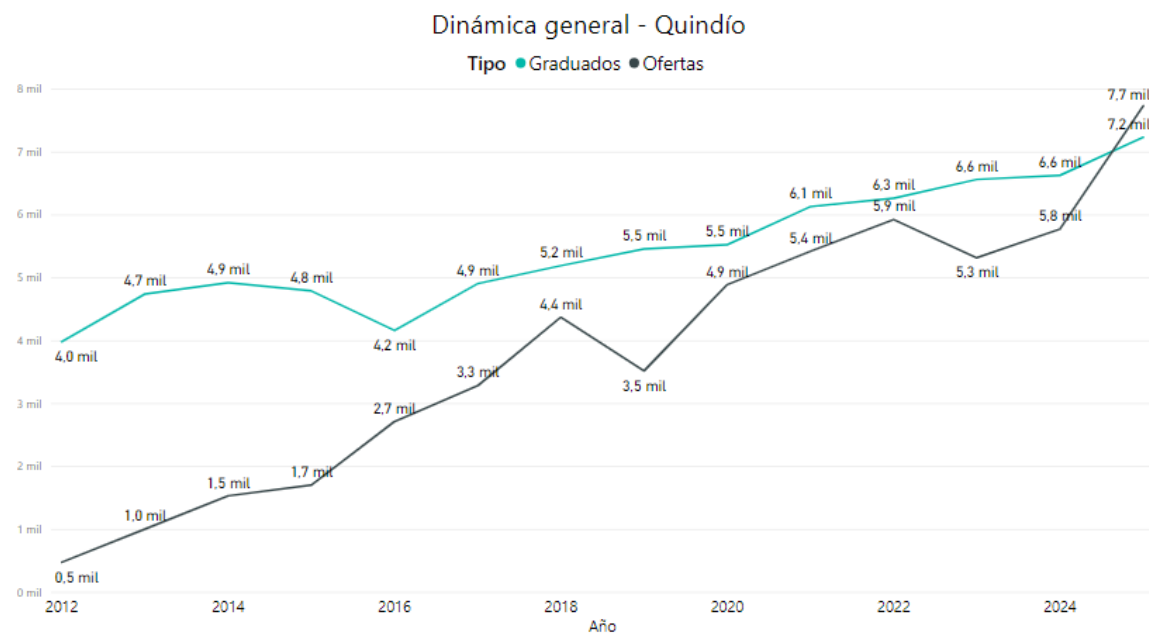


Figura 54. Dinámica de mercado general - Departamento de Quindío

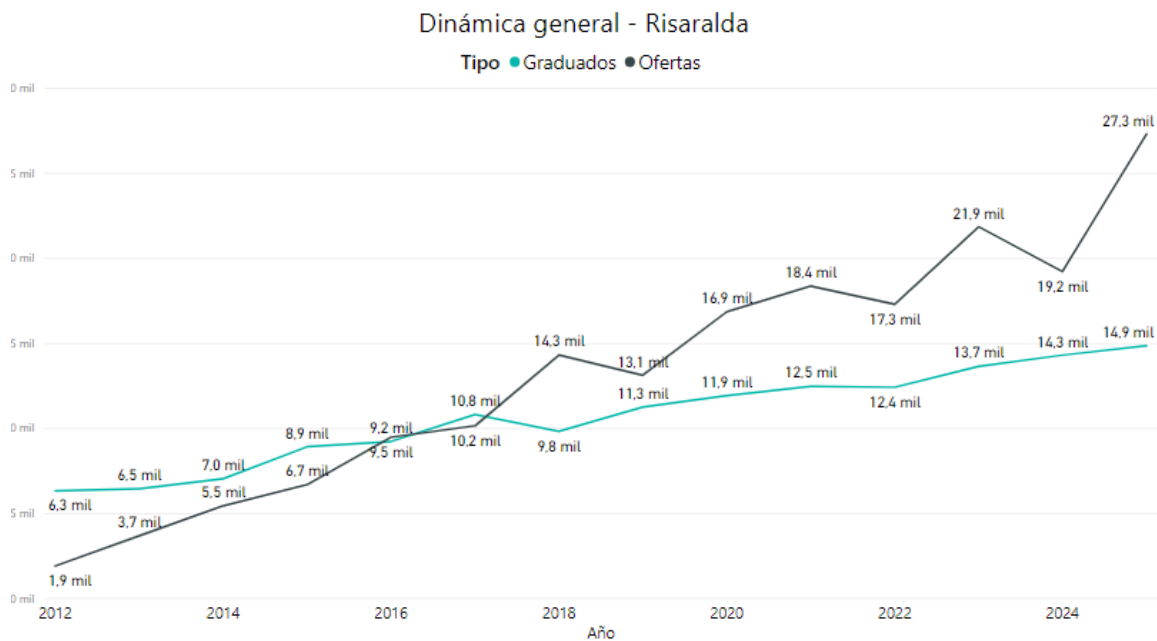


Figura 55. Dinámica de mercado general - Departamento de Risaralda

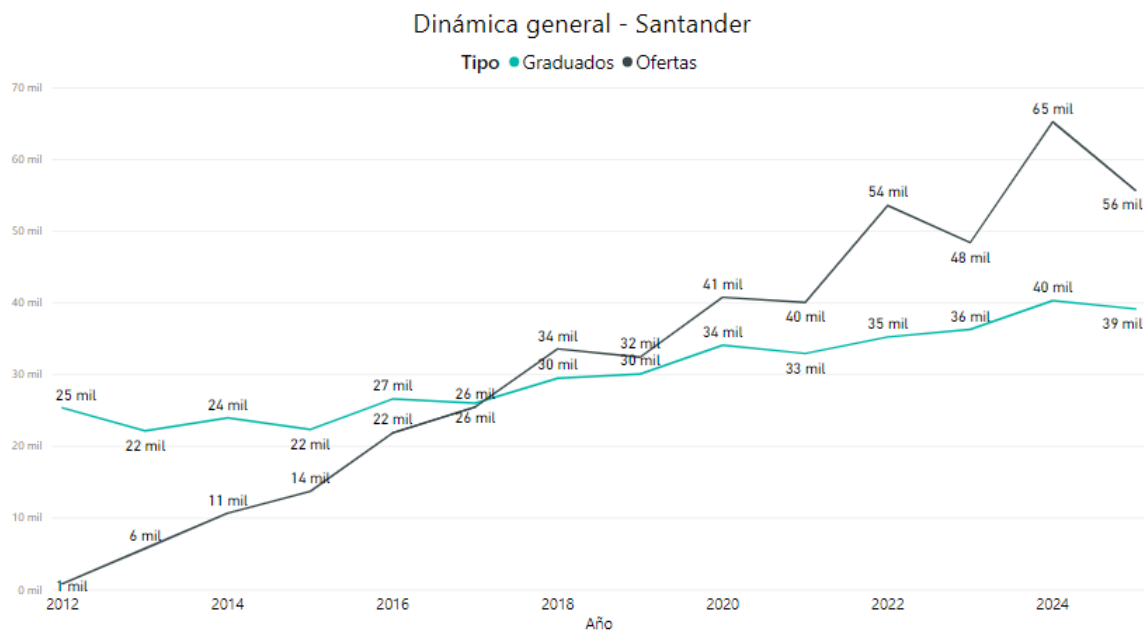


Figura 56. Dinámica de mercado general - Departamento de Santander

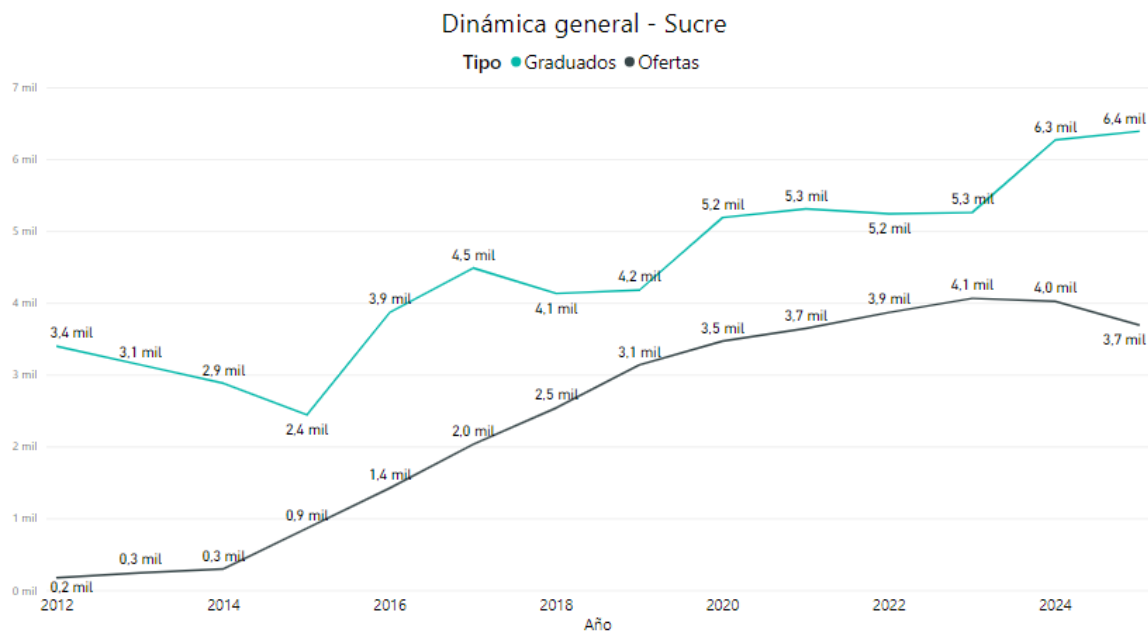


Figura 57. Dinámica de mercado general - Departamento de Sucre

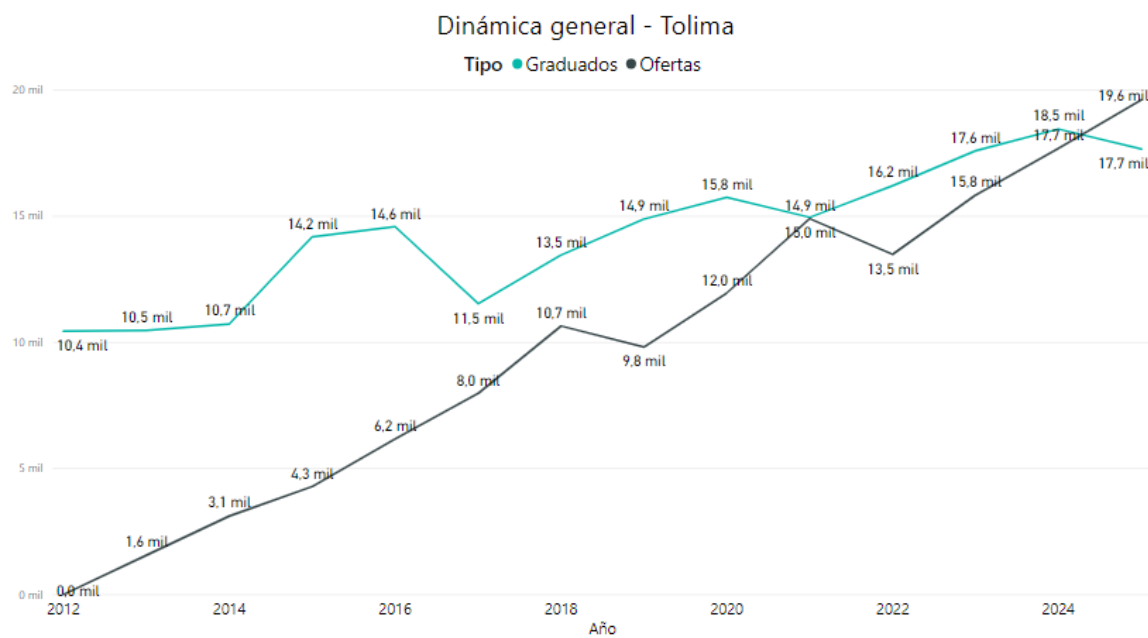


Figura 58. Dinámica de mercado general - Departamento de Tolima

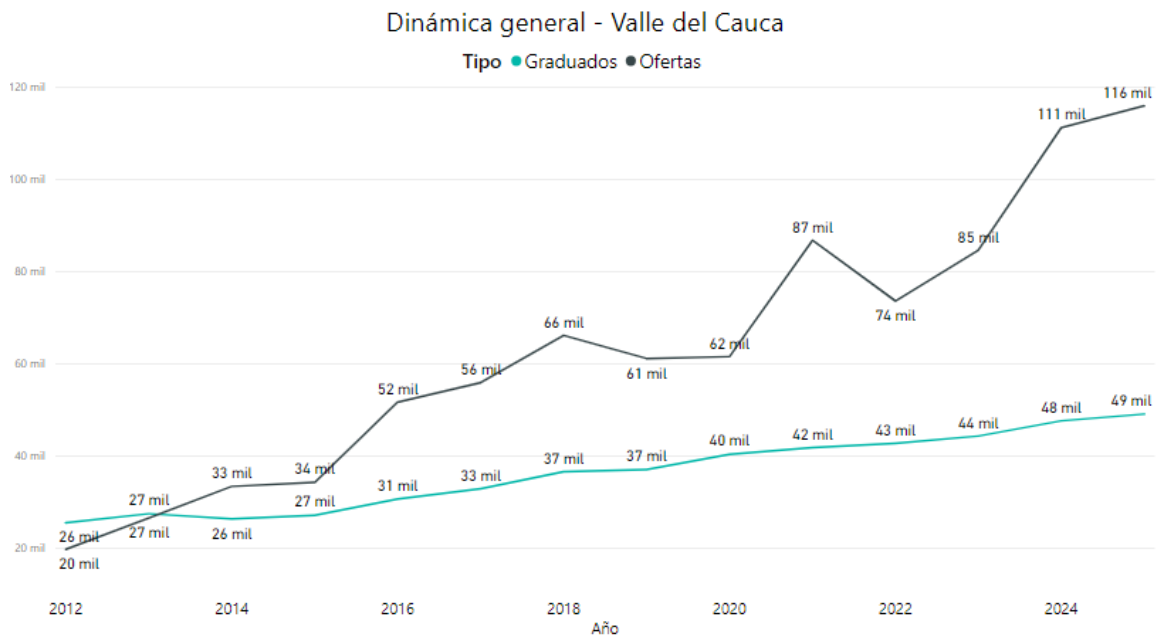


Figura 59. Dinámica de mercado general - Departamento de Valle del Cauca

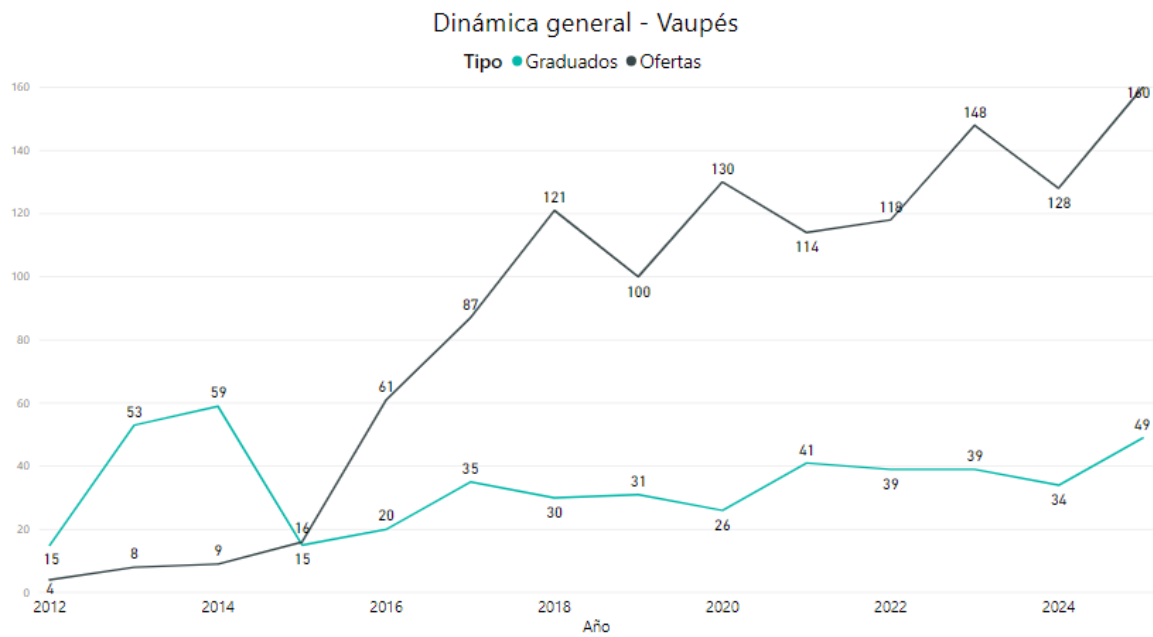


Figura 60. Dinámica de mercado general - Departamento de Vaupés

## Conclusiones

A lo largo de este trabajo de investigación se ha construido un estudio de tendencia del mercado laboral colombiano y mediante el uso de algoritmos de minería de datos describir un panorama de su comportamiento proyectando los índices de los dos componentes principales en el funcionamiento de un mercado, la oferta y la demanda, en este caso, la demanda en forma de las vacantes laborales activas en las distintas regiones del país, y la oferta, como el volumen de graduados que ofrecen las instituciones educativas a la dinámica laboral.

Las proyecciones que se generaron, sirvieron de base para identificar distintas falencias e incongruencias en el comportamiento del mercado laboral objeto de estudio, se ha evidenciado la presencia de una notable brecha, brecha que separa el número de graduados y de ofertas laborales a nivel nacional, creando lo que se decidió llamar una deficiencia generalizada de fuerza laboral en el país, sin embargo, se realizó un contraste con algunas de las cifras oficiales del país, específicamente con los índices de desempleo, y en su comparación se generó una incongruencia de resultados, se sabe que el porcentaje de desempleo actual es el más elevado en décadas, pero, en nuestra investigación encontramos un mercado saturado de ofertas laborales, por lo que se decidió tratar de explicar este fenómeno, y en el proceso, se encontró que la deficiencia se hallaba presente en algunos departamentos del país, concretamente en los tres principales, Cundinamarca, Antioquia y Valle del Cauca, pero por otro lado, en la mayoría de ciudades se identificó un comportamiento invertido, en algunos casos el mercado ofrecía más profesionales y estudiados que la capacidad de la industria de albergarlos a todos en un empleo, sin

embargo, esta información por sí sola no podría explicar este comportamiento, ya que el desempleo también estaba presente en aquellas ciudades capitales de los tres departamentos que mencionamos anteriormente, y en algunos de los casos en los que se suponía el desempleo tenía que ser mayor, se evidenciaba lo contrario, los índices de desempleo más bajos del país. Este comportamiento errático hace entonces a plantear la hipótesis ¿cumplen los profesionales del país con los requisitos de la demanda laboral del país? es una respuesta que no se puede estructurar con claridad, pero se hace énfasis en que tal vez las habilidades y conocimientos que se enseñan actualmente en un programa académico promedio, no se ajusta a la evolución que ha venido experimentando el mundo, lo que causa desequilibrios como el que hemos visto en esta investigación, y con él, la necesidad de la industria de buscar fuerza laboral extranjera que pueda suplir eficientemente sus necesidades.

La presente investigación abre las puertas a un problema investigativo mucho más profundo, hablamos de las características que un profesional tienen que adaptar a su proceso de formación y de implementación en esta era de agresiva evolución tecnológica e industrial, ¿qué factores originan el comportamiento denotado en este artículo? ¿dinámicas sociales? o una simple deficiencia de las universidades para enseñar lo que la industria necesita realmente? son hipótesis que necesitan ser estudiadas con el objetivo de explicar de manera concisa la relación entre el desempleo y el desequilibrado mercado que se ha analizado anteriormente.



## Explorador de información – anexo

EL explorador de datos que fue planteado como una parte fundamental en el desarrollo del proyecto y parte del objetivo general fue desarrollado como un informe interactivo de power bi, herramienta en la que se decidió también, moldear y graficar los resultados que se obtuvieron al implementar los algoritmos de minería seleccionados, sobre los conjuntos de datos fuente de investigación.

Power bi posee un amplio catálogo de funciones que posibilitan en este caso construir exploradores interactivos muy completos y en adición a esto, permite publicar informes en la web o en caso de necesitarlo, convertirlo en un fragmento HTML que puede ser incrustado en una página web desde su código fuente, a continuación, se observan algunas imágenes de los resultados a los que se llegaron con el explorador que se construyó:

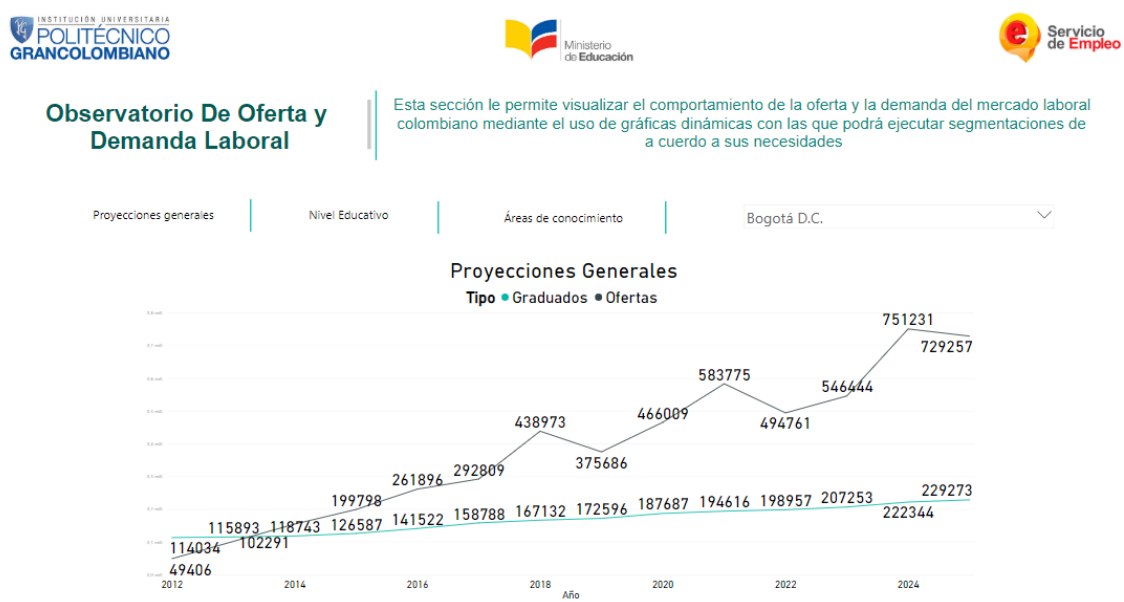


Figura 29. Pantalla principal explorador – Proyecciones generales

**Observatorio De Oferta y Demanda Laboral**

Esta sección le permite visualizar el comportamiento de la oferta y la demanda del mercado laboral colombiano mediante el uso de gráficas dinámicas con las que podrá ejecutar segmentaciones de acuerdo a sus necesidades

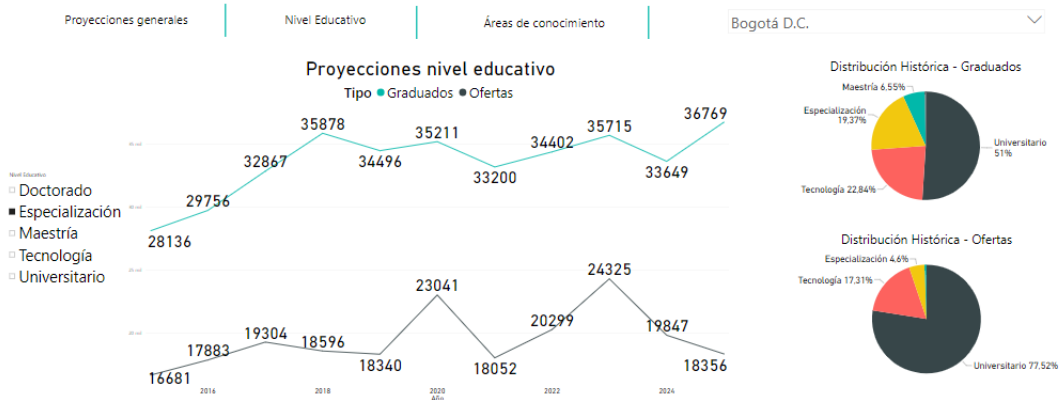


Figura 30. Pantalla proyecciones por nivel educativo

**Observatorio De Oferta y Demanda Laboral**

Esta sección le permite visualizar el comportamiento de la oferta y la demanda del mercado laboral colombiano mediante el uso de gráficas dinámicas con las que podrá ejecutar segmentaciones de acuerdo a sus necesidades

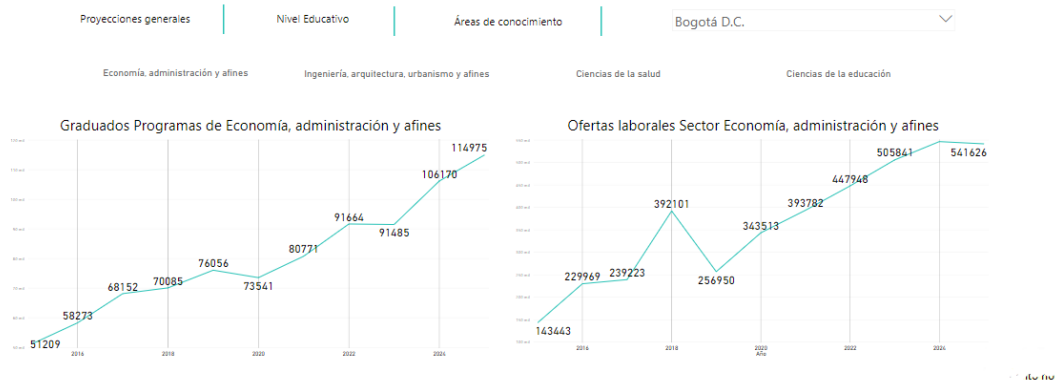


Figura 31. Pantalla proyecciones por sectores productivos

**Enlace de acceso:**

<https://app.powerbi.com/view?r=eyJrIjoiOTkwNjU1MDMtNTVIYi00ZWQzLTkwYjctNDI3NjAyMDlkYzdhIiwidCI6ImRkNTA1YmU1LWVjNjktNDdmNS05MmRmLWNhYTU1ZmViZjVmYSIsImMiOiR9>

## Bibliografía

- Alcover, R., Benlloch, J., Blesa, P., Calduch, M., & Celma, M. (2012, 16 agosto). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. Recuperado 22 septiembre, 2019, de <http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/alanal.pdf>
- Así arranca el 2019 en materia laboral. Informe de tendencias de empleo.com. (2019, 7 febrero). Recuperado 22 septiembre, 2019, de <https://www.empleo.com/co/noticias/investigacion-laboral/asi-arranca-el-2019-en-materia-laboral-informe-de-tendencias-de>
- Colombia necesita más profesionales en tecnología. (2019, 3 abril). Recuperado 22 septiembre, 2019, de <https://www.semana.com/educacion/articulo/porque-estudiar-tecnologia/607932>
- Eduardo, P., Gladys, D., & Maria, L. (2012, 16 agosto). Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios. Recuperado 22 septiembre, 2019, de [http://sedici.unlp.edu.ar/bitstream/handle/10915/19846/Documento\\_completo.pdf?sequence=1&isAllowed=y](http://sedici.unlp.edu.ar/bitstream/handle/10915/19846/Documento_completo.pdf?sequence=1&isAllowed=y)
- El desafío de elegir carrera entre la demanda laboral y lo que le gusta. (s.f.). Recuperado 22 septiembre, 2019, de <https://www.elcolombiano.com/colombia/educacion-en-colombia-harold-castilla-rector-de-corporacion-universitaria-minuto-de-dios-sobre-retos-en-educacion-ND11478027>
- Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, Lukasz Andrzej Kurgan, (2007), *Data Mining: A Knowledge Discovery Approach*, Springer Science & Business Media.
- Lanzan modelo educativo en Colombia para responder a demanda laboral. (2018, 14 diciembre). Recuperado 22 septiembre, 2019, de <https://www.empleo.com/co/noticias/noticias-laborales/lanzan-modelo-educativo-en-colombia-para-responder-demanda-laboral-5755>
- Las carreras del futuro, una elección minoritaria en Colombia. (2019, 17 septiembre). Recuperado 22 septiembre, 2019, de

<https://www.eltiempo.com/tecnosfera/novedades-tecnologia/cuantos-estudian-ciencia-ingenieria-o-tecnologia-en-colombia-412116>

Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Anuh Karpatne, (2018), Introduction to Data Mining (2nd Edition), Pearson.

¿Cuáles son las carreras con más demanda laboral en Colombia? (2019, 30 mayo). Recuperado 22 septiembre, 2019, de <https://www.semana.com/educacion/articulo/cuales-son-las-carreras-mas-demandadas-en-colombia/616638>

¿Están listos los jóvenes para el mercado laboral? (2018, 27 julio). Recuperado 22 septiembre, 2019, de <https://www.eempleo.com/co/noticias/tendencias-laborales/estan-listos-los-jovenes-para-el-mercado-laboral-5643>

¿Para qué profesionales hay más empleo en 2019? (2019, 22 junio). Recuperado 22 septiembre, 2019, de <https://www.eempleo.com/co/noticias/investigacion-laboral/para-que-profesionales-hay-mas-empleo-en-2019-5834>

Gallardo Arancibia, J. (s.f.). Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM. Recuperado de [http://www.oldemarrodriguez.com/yahoo\\_site\\_admin/assets/docs/Documento\\_CRISP-DM.2385037.pdf](http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.pdf)

Ahmed Banafa, A. (2018, 1 agosto). Small Data frente a Big Data: de vuelta a lo básico | OpenMind. Recuperado 5 enero, 2020, de <https://www.bbvaopenmind.com/tecnologia/mundo-digital/small-data-frente-a-big-data-de-vuelta-a-lo-basico/>

Estadísticas - Sistemas información. (s.f.). Recuperado 16 enero, 2020, de <https://www.mineducacion.gov.co/sistemasinfo/Informacion-a-la-mano/212400:Estadisticas>

RapidMiner GmbH. (s.f.). Operator Manual - RapidMiner Documentation. Recuperado 16 enero, 2020, de <https://docs.rapidminer.com/latest/studio/operators/>