

**ANÁLISIS DE INFORMACIÓN DE LAS BASES DE DATOS DE LOS ESTUDIANTES
DEL POLITÉCNICO GRANCOLOMBIANO CON EL OBJETIVO DE GENERAR UN
PERFIL DE INGRESO**

AUTOR:

JORGE ANDRÉS ALDANA SILVA – CÓDIGO 1510650247

ASESOR: MSC. GIOVANNY ALEXANDER BAQUERO VILLAMIL

**INSTITUCIÓN UNIVERSITARIA POLITÉCNICO GRANCOLOMBIANO
FACULTAD DE INGENIERÍA, DISEÑO E INNOVACIÓN
ESCUELA DE OPTIMIZACIÓN, PRODUCCIÓN, INFRAESTRUCTURA Y
AUTOMATIZACIÓN
ESPECIALIZACIÓN EN GERENCIA DE PROYECTOS EN INTELIGENCIA DE
NEGOCIOS
BOGOTÁ, D.C. 2019**

TABLA DE CONTENIDO

	Pág.
1. TÍTULO.....	5
2. RESUMEN	5
2.1 ABSTRACT.....	5
3. TEMA.....	6
4. FUNDAMENTACIÓN DEL PROYECTO.....	6
4.1 MARCO CONTEXTUAL	6
5. PROBLEMA	9
6. JUSTIFICACIÓN.....	9
7. OBJETIVO GENERAL	9
8. MARCO CONCEPTUAL.....	9
9. OBJETIVOS ESPECÍFICOS, ACTIVIDADES Y CRONOGRAMA.....	11
9.1 OBJETIVOS ESPECÍFICOS.....	11
9.2 METODOLOGÍA	11
10. ANÁLISIS	13
10.1 ANÁLISIS CORRELACIÓN NOTAS VS ASESORÍAS	13
10.2 ANÁLISIS EMPRESAS DONDE TRABAJAN LOS ESTUDIANTES.....	19
11. OBSERVACIONES	20
12. RECOMENDACIONES.....	21

TABLA DE FIGURAS

	Pág.
Figura 1. Cantidad de estudiantes matriculados por periodo.....	7
Figura 2. Cantidad de estudiantes matriculados por jornada en modalidad presencial.	8
Figura 3. Estudiantes matriculados por periodo y nivel de formación.	8

TABLA DE CUADROS

	Pág.
Cuadro 1. Descripción de la tabla a analizar.....	12
Cuadro 2. Fórmulas para análisis descriptivo.	13
Cuadro 3. Análisis descriptivo 1.....	14
Cuadro 4. Matriz de correlaciones 1.	15
Cuadro 5. Análisis descriptivo 2.....	16
Cuadro 6. Matriz de correlaciones 2.	16
Cuadro 7. Análisis descriptivo 3.....	17
Cuadro 8. Matriz de correlaciones 3.	17
Cuadro 9. Análisis descriptivo 4.....	17
Cuadro 10. Matriz de correlaciones 4.	18
Cuadro 11. Análisis descriptivo 5.....	18
Cuadro 12. Matriz de correlaciones 5.	19
Cuadro 13. Nombres con que se registra la entidad Policía Nacional.	19

1. TÍTULO

Análisis de información de las bases de datos de los estudiantes del politécnico grancolombiano con el objetivo de generar un perfil de ingreso, y análisis de las bases de datos de las notas de los estudiantes de modalidad presencial y asesorías extra clase para encontrar la incidencia de las asesorías con el comportamiento de la nota.

2. RESUMEN

La información se ha convertido en un activo muy importante para todas las empresas, ya que a partir de ésta se pueden generar varios análisis orientados al conocimiento de negocio, y si se cuenta con información de buena calidad, se pueden hacer análisis más profundos orientados al apoyo en la toma de decisiones.

Las empresas tienen normalmente un ERP en donde registran casi toda su actividad comercial, y cuentan con otros sistemas diseñados propiamente en la empresa en donde se registran actividades propias del negocio o de la empresa que no están incluidas en el ERP.

El Politécnico Grancolombiano cuenta con un sistema de información llamado NEWCIO, en donde registra la información correspondiente a los estudiantes y sus actividades, además cuenta con un sistema aparte diseñado por el área de informática en donde se registran otras actividades de los estudiantes como las asesorías extra clase.

Este proyecto tiene como fin tomar la información de NEWCIO y la información de las asesorías extra clase para encontrar la relación existente entre la asistencia a las asesorías y las notas obtenidas durante el semestre.

También se puede, a partir de un análisis, determinar las empresas o colegios de los cuales viene la mayoría de estudiantes para ofrecer algunos beneficios para los nuevos estudiantes que provengan de esas empresas o colegios.

2.1 ABSTRACT

Information has become a very important asset for all companies, since from it can generate several analyzes oriented to business knowledge, and if you have good quality information, you can make more in-depth analysis aimed at Support in decision making.

Companies usually have an ERP where they register almost all their commercial activity, and have other systems designed properly in the company where they register activities of the business or the company that are not included in the ERP.

The Politécnico Grancolombiano has an information system called NEWCIO, where it records the information corresponding to students and their activities, also has a separate system designed by the computer area where other student activities such as extra class consultancies are recorded.

The purpose of this project is to take the information from NEWCIO and the information from the extra class consultancies to find the relationship between the assistance to the consultancies and the grades obtained during the semester.

You can also, from an analysis, determine the companies or schools from which most students come to offer some benefits for new students coming from these companies or schools.

3. TEMA

El Politécnico Grancolombiano es una Institución Universitaria privada, de utilidad común y como tal carece de fines de lucro. Tiene como domicilio la ciudad de Bogotá, Distrito Capital, República de Colombia, pero puede establecer dependencias, seccionales u otros centros de actividad en cualquier lugar del país o del exterior.

El objeto de la Institución Universitaria Politécnico Grancolombiano (IUPG) es la prestación de servicios en todos los campos de acción de la educación, en especial en el de la educación superior, según lo previsto en el Artículo 7 de la Ley 30 de 1992 y demás disposiciones complementarias y concordantes.

El Politécnico Grancolombiano tiene un departamento de inteligencia competitiva en donde se desarrollan estrategias que permiten aumentar la cantidad de estudiantes matriculados por periodo, analizar las necesidades del mercado para ofrecer nuevos programas de formación y establecer convenios con entidades colegios, instituciones y empresas para dar beneficios a los estudiantes provenientes de allí.

4. FUNDAMENTACIÓN DEL PROYECTO

4.1 MARCO CONTEXTUAL

El Politécnico Grancolombiano es una Institución Universitaria privada, de utilidad común y como tal carece de fines de lucro. Tiene como domicilio la ciudad de Bogotá, Distrito Capital, República de Colombia, pero puede establecer dependencias, seccionales u otros centros de actividad en cualquier lugar del país o del exterior.

El objeto de la Institución Universitaria Politécnico Grancolombiano (IUPG) es la prestación de servicios en todos los campos de acción de la educación, en especial en el de la educación superior, según lo previsto en el Artículo 7 de la Ley 30 de 1992 y demás disposiciones complementarias y concordantes.

Misión

La Institución Universitaria Politécnico Grancolombiano tiene como fin contribuir a la inclusión social y al desarrollo de la nación, a través de programas en toda la cadena de formación que se distinguen por su calidad y pertinencia y el desarrollo de proyectos de extensión e investigación

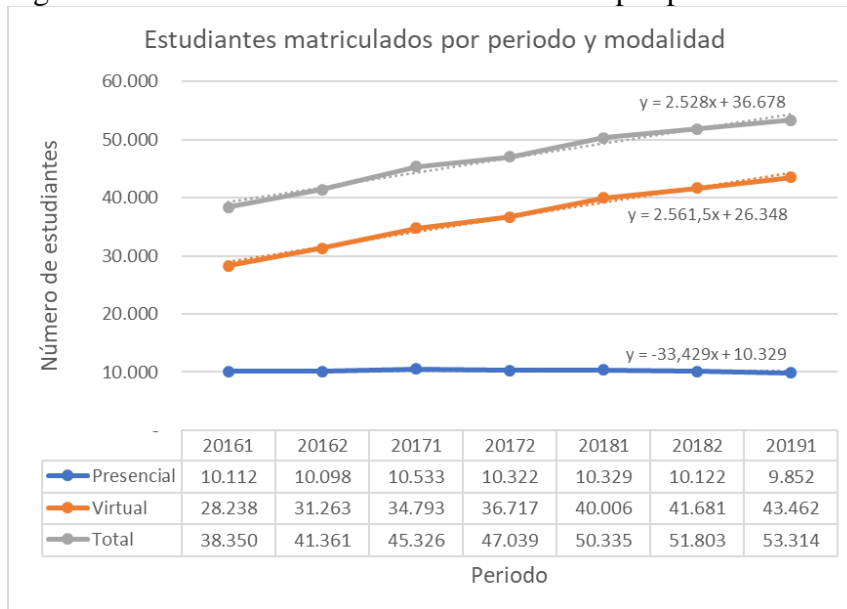
aplicada, fundamentados en los valores institucionales con el firme propósito de buscar la excelencia académica.

Visión 2013-2017

Para el 2017, la IUPG será reconocida en el ámbito nacional e internacional como la Institución de Educación Superior que brinda más y mejores alternativas educativas en todos los niveles de la cadena de formación, por su compromiso con la calidad, la pertinencia, la inclusión social y el desarrollo de proyectos con incidencia en el sector social y empresarial.

El Politécnico Grancolombiano cuenta actualmente con 53.314 estudiantes matriculados, 9.852 en modalidad presencial en 45 programas diferentes y 43.462 en modalidad virtual en 50 programas diferentes, la cantidad de estudiantes matriculados ha aumentado desde el periodo 20161 en la modalidad virtual en 2.561 por periodo y en modalidad presencial se ha mantenido estable la cantidad de estudiantes matriculados por periodo, excepto en el periodo 20191 donde la cantidad de estudiantes disminuyó a 9.852.

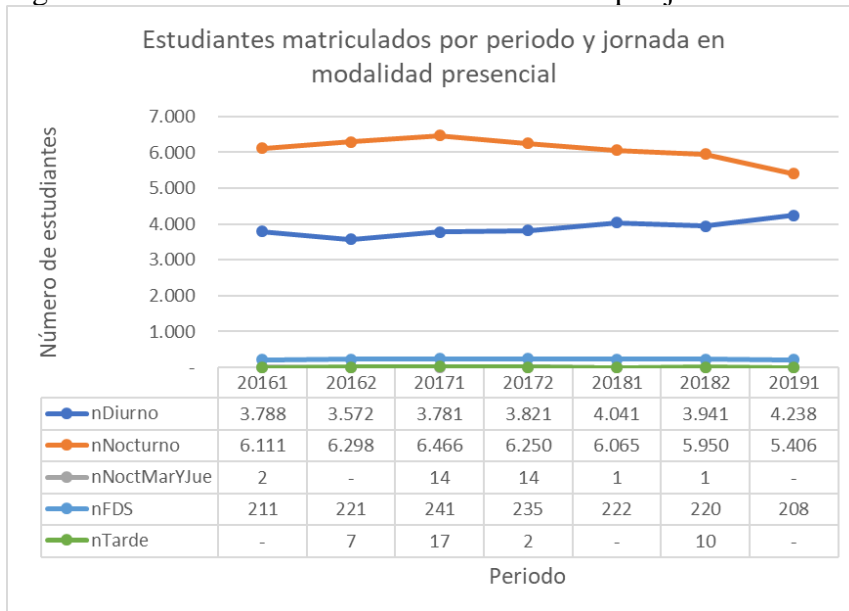
Figura 1. Cantidad de estudiantes matriculados por periodo.



Fuente: Autor.

El Politécnico Grancolombiano ofrece sus diferentes programas en jornada diurno, tarde, nocturno y fines de semana en modalidad presencial, la cantidad de estudiantes matriculados por periodo en cada jornada se muestra en la siguiente gráfica.

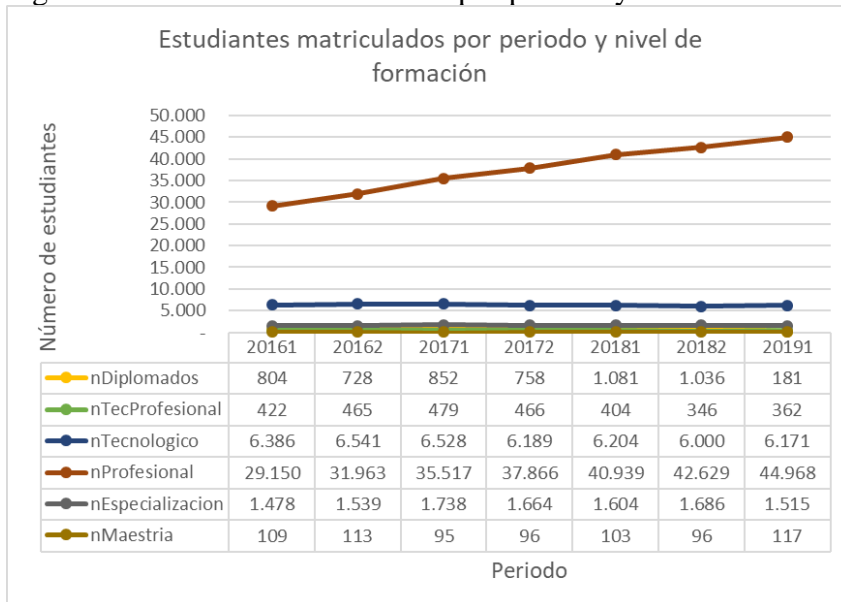
Figura 2. Cantidad de estudiantes matriculados por jornada en modalidad presencial.



Fuente: Autor.

Se ofertan programas en diferentes niveles de formación (Cursos, Seminarios, Diplomados, Técnicos Laboral, Técnico Profesional, Tecnológico, Profesional, Especialización y Maestría), en la siguiente gráfica se muestra la cantidad de matriculados por los niveles de formación más representativos.

Figura 3. Estudiantes matriculados por periodo y nivel de formación.



Fuente: Autor.

5. PROBLEMA

El Politécnico Grancolombiano cuenta con muchas herramientas (asesorías extra clase, talleres, etc.) para ayudar a los estudiantes a reforzar sus conocimientos en algunas materias, pero no tiene la certeza de qué tan eficientes son estas herramientas para lograr el objetivo.

El desconocimiento sobre lo que hace que las herramientas diseñadas sean o no efectivas puede provocar que no se haga un buen uso de los recursos existentes, o que se esté tratando de usar una herramienta de diferentes maneras, cuando la falla tal vez no esté en la herramienta.

6. JUSTIFICACIÓN

El Politécnico Grancolombiano recibe aproximadamente 21.703 estudiantes nuevos al año, de los cuales solo el 72,43% se vuelve a matricular para segundo semestre.

Las causas por las que los estudiantes no continúan en el Politécnico Grancolombiano son muchas, entre ellas están factores económicos, factores familiares, bajo desempeño en el semestre, entre otras.

Este proyecto busca explorar la causa de desempeño académico en el semestre, y para ello se tomará información de los estudiantes, sus notas y las asistencias a las asesorías extra clase.

Se va a analizar qué tanto inciden las notas del estudiante para que éste tome la decisión de no continuar en la universidad y también se va a analizar qué tanto inciden las asesorías extra clase en el desempeño académico del estudiante. Esto se hace para ver qué tanto riesgo hay de que un estudiante no siga en la universidad por motivo académico y qué tan efectivas son las asesorías extra clase.

Con este análisis se puede diseñar, posteriormente, un plan de acción para motivar a los estudiantes a asistir a las asesorías extra clase para que mejoren su desempeño académico.

7. OBJETIVO GENERAL

Analizar la información de las tablas de matriculados, notas, asesorías extra clase y calendario académico, para encontrar relaciones entre diferentes variables.

8. MARCO CONCEPTUAL

Para el desarrollo del proyecto se usarán varias herramientas tecnológicas y conceptos de manejo de información y estadísticos.

Sistema manejador de bases de datos (DBMS – DataBase Management System)

Un **sistema manejador de bases de datos** o DBMS es una colección de datos interrelacionados y un grupo de programas para acceder a esos datos. La colección de datos se llama comúnmente *base de datos*, contiene información relevante de una empresa. El principal objetivo de un DBMS es proporcionar una manera de almacenar y devolver información de una base de datos de una manera conveniente y eficiente.

Los sistemas de bases de datos están diseñados para administrar grandes volúmenes de información. La administración de datos involucra la definición de las estructuras para almacenamiento de información y proporcionar mecanismos para manipular la información. Además, el sistema de base de datos debe garantizar la seguridad de la información almacenada, a pesar de las fallas del sistema o intentos de acceso no autorizado. Si los datos deben ser compartidos entre varios usuarios, el sistema debe evitar posibles resultados anómalos.

Los sistemas manejadores de bases de datos más conocidos y utilizados son SQL Server (con licencia), Oracle (con licencia), PostgreSQL (libre), MySQL (con licencia), MS Access (con licencia).

Para poder acceder a la información de una base de datos, se usa generalmente código **SQL** (Structured Query Language), que es un lenguaje de consulta estructurado. Con este lenguaje se puede ver información de una o varias tablas con cálculos, filtros, transformaciones, entre otros.

El lenguaje SQL más básico es de la forma SELECT-FROM-WHERE, en donde SELECT indica el campo o campos que se quieren ver, la sentencia FROM indica la tabla o conjunto de tablas de donde se va a sacar la información y la sentencia WHERE indica las condiciones de dicha información.

Por ejemplo, se puede ver la información de todos los estudiantes de modalidad presencial con el siguiente código:

```
SELECT *  
FROM T_Estudiantes  
WHERE modalidad = 'Presencial'
```

La información extraída de un sistema de bases de datos se puede utilizar para ser analizada y tener más información acerca del negocio, existen métodos estadísticos que permiten realizar estos análisis.

El método más sencillo para el análisis inicial de los datos es un **análisis descriptivo multivariante** que tomando variables numéricas, muestra la naturaleza de los datos y llega hasta el análisis de correlaciones, que es ver qué tanto se puede representar una variable por medio de otra, es decir, la relación existente entre las variables.

Las medidas básicas descriptivas son la media, el rango, la varianza, la desviación estándar y el coeficiente de variación.

Hay medidas descriptivas más avanzadas como la curtosis, el coeficiente de asimetría y el coeficiente de correlación.

La **media** de un conjunto de datos es la suma de los valores de los datos dividida por el número de observaciones.

El **rango** es la diferencia entre la observación mayor (valor máximo) y la menor (valor mínimo).

la **varianza muestral**, s^2 , es la suma de los cuadrados de las diferencias entre cada observación y la media muestral dividida por el tamaño de la muestra, n , menos 1.

La **desviación estándar** es la raíz cuadrada de la varianza.

El **coeficiente de variación**, CV, es una medida de la dispersión relativa que expresa la desviación típica en porcentaje de la media (siempre que la media sea positiva).

El **coeficiente de correlación**, r , muestra qué tanto puede una variable representar a otra, este coeficiente varía puede tomar valores en el rango $[-1, 1]$, si la correlación tiene valor -1, indica una relación perfecta entre dos variables en donde, mientras una crece, la otra decrece, y si la correlación es 1, indica una relación perfecta en donde cuando una variable crece, la otra también crece.

9. OBJETIVOS ESPECÍFICOS, ACTIVIDADES Y CRONOGRAMA

9.1 OBJETIVOS ESPECÍFICOS

- Determinar la incidencia de las asesorías extra clase en las notas de los cortes de los estudiantes.
- Encontrar las empresas de las cuáles llegan más estudiantes nuevos al Politécnico Grancolombiano.
- Generar un perfil de ingreso de los estudiantes nuevos.

9.2 METODOLOGÍA

Para realizar el análisis de datos de las bases de datos del Politécnico Grancolombiano, se cargan las tablas recibidas en archivos Excel a una base de datos en SQL y se utiliza lenguaje SQL (Structures Query Language) para organizar la información como se necesita para realizar un análisis estadístico descriptivo.

A cada tabla cargada se le va a generar una vista, esta vista se encarga de convertir varios registros en uno solo. Eg. La tabla de notas muestra un registro por la nota de cada cohorte, es decir, la tabla muestra 4 registros por estudiante, materia y periodo. La vista generada para esta tabla muestra ahora un registro por estudiante y materia, y en cuatro columnas muestra las diferentes notas.

Luego los resultados de las vistas se insertan en nuevas tablas, esto se hace porque algunas vistas son muy pesadas para ejecutarse lo cual retrasaría el análisis de los datos, y con estas nuevas tablas se harán más vistas.

Se hará unión de las tablas de matriculados, notas y asesorías, todo esto de la modalidad presencial.

También se añadirán campos calculados como la pendiente de la recta generada por los cohortes 1, 2 y 3, y las notas obtenidas, y la edad de los estudiantes.

A continuación se muestran los campos que serán analizados.

Cuadro 1. Descripción de la tabla a analizar.

Variable	Tabla/Vista	Comentarios
nota1	tmpNotasPresencial	
nota2	tmpNotasPresencial	
nota3	tmpNotasPresencial	
nota4	tmpNotasPresencial	
fallas1	tmpNotasPresencial	
fallas2	tmpNotasPresencial	
fallas3	tmpNotasPresencial	
fallas4	tmpNotasPresencial	
nTalleres	tmpAsesoriasPresencial	
nAsesorias	tmpAsesoriasPresencial	
nTalleresMasAsesorias	tmpAsesoriasPresencial	
edad	T_Matricula2	Campo calculado
correlxnota		Campo calculado: coeficiente de correlación entre los cohortes 1, 2 y 3, y las notas obtenidas.
m		Campo calculado: pendiente de la recta de regresión lineal calculada con los cohortes 1, 2 y 3 (en x), y las notas obtenidas (en y) para cada estudiante.

Fuente: Autor.

Con los campos mostrados anteriormente y varios filtros que serán aplicados durante el análisis, se realizará un análisis estadístico descriptivo Excel con operaciones matriciales y con técnicas de análisis multivariante.

A continuación se muestran las fórmulas que serán utilizadas en dicho análisis.

Cuadro 2. Fórmulas para análisis descriptivo.

Medida	Forma algebraica	Forma matricial	Interpretación
Media (vector de medias)	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{X}_{p \times 1} = \frac{1}{n} X_{p \times n}^T \mathbf{1}_{n \times 1}$	
Matriz de datos centrada		$\tilde{X}_{n \times p} = X_{n \times p} - \mathbf{1}_{n \times 1} \bar{X}_{1 \times p}^T$	
Varianza poblacional (matriz de varianzas)	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$	$S_{p \times p} = \frac{1}{n} \tilde{X}_{p \times n}^T \tilde{X}_{n \times p}$	
Varianza muestral (Matriz de varianza corregida)	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$	$\tilde{S}_{p \times p} = \frac{1}{n - 1} \tilde{X}_{p \times n}^T \tilde{X}_{n \times p}$	
Matriz de correlación	$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$	$R_{p \times p} = D^{-\frac{1}{2}} \tilde{S} D^{-\frac{1}{2}}$	
Coefficiente de variación (CV)	$CV = \frac{s}{\bar{x}}$		
Coefficiente de asimetría (A)	$CA = \frac{1}{n} \frac{\sum (x_i - \bar{x})^3}{s^3}$		Es cero para una distribución simétrica, si su abs es mayor que uno es asimétrico
Coefficiente de curtosis (K)	$K = \frac{1}{n} \frac{\sum (x_i - \bar{x})^4}{s^4}$		Valor mayor a 7 u 8 si hay heterogeneidad por pocos valores atípicos alejados del resto, si está entre 1 y 2 indica mezcla entre dos distribuciones.

Fuente: Autor.

Se van a utilizar las fórmulas mostradas en forma matricial en la tabla. Esto facilita los cálculos y muestra de una manera sencilla el comportamiento de los datos.

10. ANÁLISIS

Se reciben 3 tablas desde el área de permanencia para realizar un análisis a partir de ellas.

Una tabla contiene los datos de los estudiantes matriculados desde el periodo 20161 hasta el periodo 20191, otra tabla contiene las notas de los estudiantes de la modalidad presencial desde el periodo 20171 hasta el periodo 20191 y la otra tabla contiene los datos de las asesorías extra clase en modalidad presencial desde el periodo 20171 hasta el periodo 20182.

Las tablas se reciben en archivos Excel y se cargan en SQL Server Express Edition 2014 para realizar las vistas y cruces necesarios entre ellas para facilitar el análisis de la información.

10.1 ANÁLISIS CORRELACIÓN NOTAS VS ASESORÍAS

Se va a hacer un análisis descriptivo multivariante a los campos nota1, nota2, nota3, nota4, fallas1, fallas2, fallas3, fallas4, nTalleres, nAsesorías, nTalleresMasAsesorias, edad, correlxnota y m. La descripción de estos datos está en el cuadro 1.

Inicialmente se creó una vista con los campos mencionados anteriormente, en donde se muestran 146.736 registros.

La idea del análisis descriptivo multivariante es mostrar algunas medidas de las variables como el valor mayor, el valor menor, el promedio, la varianza (muestral), la desviación estándar (muestral), coeficiente de variabilidad, coeficiente de asimetría, coeficiente de homogeneidad y curtosis.

Con este análisis, se puede comprender los datos inicialmente, el análisis puede realizarse estandarizando o sin estandarizar las variables, para este trabajo, se hará el análisis sin estandarizar las variables.

El análisis descriptivo también permite ver si los datos que se van a estudiar son coherentes o no, por ejemplo, al realizar el primer análisis, se encontraron 354 estudiantes a los que la edad calculada aparece con valores entre 0 y 15, y aparecen también 18 estudiantes que no tienen fecha de nacimiento registrada. Estos registros se omiten para el análisis.

A continuación se muestra el cuadro con el análisis descriptivo multivariante realizado.

Cuadro 3. Análisis descriptivo 1.

Descripción de los datos														
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas3	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
Máximo	5,00	5,00	5,00	5,00	11,00	8,00	9,00	26,00	7,00	26,00	26,00	67,83	1,00	2,50
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	16,02	-1,00	-2,50
Rango	5,00	5,00	5,00	5,00	11,00	8,00	9,00	26,00	7,00	26,00	26,00	51,81	2,00	5,00
\bar{X}	3,63	3,62	3,68	3,70	0,00	0,00	0,00	0,01	0,01	0,06	0,06	23,67	0,08	0,02
Varianza	0,86	1,04	1,24	0,68	0,01	0,00	0,00	0,07	0,02	0,19	0,22	30,72	0,53	0,33
Desviación	0,93	1,02	1,12	0,82	0,10	0,07	0,05	0,26	0,14	0,44	0,47	5,54	0,73	0,58
C. Variab.	0,2556	0,2814	0,3030	0,2226	39,3648	41,0584	69,6779	35,8294	19,7332	7,7864	7,5010	0,2341	9,1508	23,3076
C. Asimetri	-1,2987	-1,5044	-1,7034	-1,7375	57,3572	56,1435	105,7589	52,4139	27,5289	15,4672	14,7672	2,1232	-0,1750	-0,3498
C. Homog.	4,5190	4,8963	5,1807	6,7650	4.120,6559	3.979,2612	14.359,4656	3.367,5566	930,5417	390,6965	359,2081	8,4798	0,4663	5,3049
C. Curtosis	5,5190	5,8963	6,1807	7,7650	4.121,6559	3.980,2612	14.360,4656	3.368,5566	931,5417	391,6965	360,2081	9,4798	1,4663	6,3049

Fuente: Autor.

Con el análisis del cuadro anterior, se observa la coherencia de los datos, como el hecho de ver las notas entre 0 y 5, la edad máxima y mínima, coeficientes de correlación con valores entre -1 y 1, y pendiente con valores entre -2,5 y 2,5.

Con los datos útiles para el análisis, se realiza una matriz de correlaciones para ver si existen relaciones entre variables numéricas. Se espera encontrar correlaciones entre la asistencia a asesorías extra clase y pendiente de la recta de regresión lineal entre cohortes y notas.

A continuación se muestra la matriz de correlaciones

Cuadro 4. Matriz de correlaciones 1.

Matriz de correlaciones (R)														
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas3	fallas4	nTalleres	nAsesorias	nTalleresMasAsesorias	edad	correlxnota	m
nota1	1,0000	0,4354	0,3779	0,6096	-0,0172	-0,0158	-0,0075	-0,0181	-0,0276	-0,0178	-0,0244	-0,0253	-0,3242	-0,4405
nota2	0,4354	1,0000	0,5275	0,7363	-0,0248	-0,0297	-0,0154	-0,0298	-0,0245	-0,0103	-0,0166	-0,0122	0,1444	0,1602
nota3	0,3779	0,5275	1,0000	0,8361	-0,0217	-0,0264	-0,0144	-0,0270	-0,0097	0,0063	0,0030	0,0038	0,5324	0,6647
nota4	0,6096	0,7363	0,8361	1,0000	-0,0271	-0,0317	-0,0178	-0,0332	-0,0246	-0,0045	-0,0112	0,0011	0,2496	0,3188
fallas1	-0,0172	-0,0248	-0,0217	-0,0271	1,0000	0,5653	0,3355	0,8327	-0,0013	-0,0010	-0,0013	0,0302	-0,0053	-0,0072
fallas2	-0,0158	-0,0297	-0,0264	-0,0317	0,5653	1,0000	0,3803	0,8356	-0,0012	-0,0011	-0,0014	0,0268	-0,0113	-0,0128
fallas3	-0,0075	-0,0154	-0,0144	-0,0178	0,3355	0,3803	1,0000	0,5520	-0,0007	-0,0009	-0,0011	0,0184	-0,0075	-0,0079
fallas4	-0,0181	-0,0298	-0,0270	-0,0332	0,8327	0,8356	0,5520	1,0000	-0,0014	-0,0013	-0,0016	0,0311	-0,0100	-0,0115
nTalleres	-0,0276	-0,0245	-0,0097	-0,0246	-0,0013	-0,0012	-0,0007	-0,0014	1,0000	0,1175	0,3958	-0,0092	0,0125	0,0129
nAsesorias	-0,0178	-0,0103	0,0063	-0,0045	-0,0010	-0,0011	-0,0009	-0,0013	0,1175	1,0000	0,9585	-0,0271	0,0154	0,0205
nTalleresM	-0,0244	-0,0166	0,0030	-0,0112	-0,0013	-0,0014	-0,0011	-0,0016	0,3958	0,9585	1,0000	-0,0277	0,0178	0,0226
edad	-0,0253	-0,0122	0,0038	0,0011	0,0302	0,0268	0,0184	0,0311	-0,0092	-0,0271	-0,0277	1,0000	0,0181	0,0241
correlxnota	-0,3242	0,1444	0,5324	0,2496	-0,0053	-0,0113	-0,0075	-0,0100	0,0125	0,0154	0,0178	0,0181	1,0000	0,7779
m	-0,4405	0,1602	0,6647	0,3188	-0,0072	-0,0128	-0,0079	-0,0115	0,0129	0,0205	0,0226	0,0241	0,7779	1,0000

Fuente: Autor.

La matriz de correlaciones muestra la relación existente entre la nota definitiva (nota4) y las notas de cada cohorte, es de esperarse dado que la nota definitiva es casi el promedio de las notas de los cohortes. También se observa una correlación superior al 0,37 para las notas de los cohortes 1, 2 y 3, es decir, que los estudiantes tienden a sacar notas similares en los diferentes cohortes, y la correlación tan alta de nota3 con nota4 puede deberse a que nota3 representa el 40% de nota4.

Se observa también que las fallas totales (fallas4) están muy relacionadas con las fallas de los cohortes, es de esperarse porque las fallas4 tienden a ser la sumatoria de las fallas1, fallas2 y fallas3. También se observa por los coeficientes de correlación superiores a 0,33 que los estudiantes que fallan en un cohorte, tienden a fallar en los demás cohortes.

En cuanto a nTalleres, nAsesorias y nTalleresMasAsesorias, se observa que de los estudiantes que asisten a talleres, pocos asisten a asesorías extra clase, y por la correlación entre nAsesorias y nTalleresMasAsesorias, se observa que los estudiantes prefieren las asesorías a los talleres.

En cuanto a la edad, se observa que ésta no tiene correlación con ninguna variable.

Se observa correlación entre correlxnota y m, pero es de esperarse dado que un coeficiente de correlación positivo indica una pendiente positiva, y un coeficiente de correlación negativo indica pendiente negativa.

La relación más esperada, que es la relación entre nAsesorias y m, es demasiado baja, cercana a cero, pero esto puede deberse a que se tomaron los registros de todas las materias, así el estudiante no haya asistido a asesorías.

Finalmente se encuentra una relación que no se esperaba en el análisis, y es la relación entre nota3 y m, esto se debe a que muchos estudiantes no tienen registradas las notas de los cohortes 1 y 2, solo tienen registrada la nota del cohorte 3 y la definitiva.

Con el fin de reducir el grupo de estudio para hacer un grupo más específico de análisis, se hará un nuevo análisis descriptivo pero ahora teniendo en cuenta solo las notas de las materias 'Algebra abstracta I', 'Algebra abstracta II', 'Algebra Lineal', 'Cálculo I', 'Cálculo II', 'Cálculo III', 'Matemáticas', 'Matemáticas Financieras', 'Matemáticas II', 'Probabilidad', 'Estadística', 'Estadística

Básica', 'Estadística I', 'Estadística II', 'Estadística Inferencial', 'Estadística Inferencial'. El número de registros a analizar disminuye a 14.246.

A continuación se muestra el cuadro con el análisis descriptivo realizado.

Cuadro 5. Análisis descriptivo 2.

Descripción de los datos														
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas3	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
Máximo	5,00	5,00	5,00	5,00	10,00	2,00	2,00	4,00	7,00	26,00	26,00	62,89	1,00	2,50
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	16,08	-1,00	-2,50
Rango	5,00	5,00	5,00	5,00	10,00	2,00	2,00	4,00	7,00	26,00	26,00	46,81	2,00	5,00
\bar{X}_{xx}	3,25	3,16	3,25	3,28	0,00	0,00	0,00	0,00	0,06	0,15	0,21	23,08	0,03	0,00
Varianza	0,82	1,14	1,37	0,76	0,01	0,00	0,00	0,00	0,17	0,60	0,89	21,04	0,53	0,32
Desviación	0,91	1,07	1,17	0,87	0,09	0,03	0,02	0,07	0,42	0,77	0,94	4,59	0,73	0,57
C. Variab.	0,2790	0,3378	0,3597	0,2664	72,6212	51,1449	84,3949	39,7744	6,5848	5,1083	4,3968	0,1987	22,5683	-512,7345
C. Asimetri	-0,5833	-0,9540	-1,2137	-1,3449	108,5708	57,4362	84,3712	47,5208	8,9894	10,5546	8,6535	1,8754	-0,0787	-0,3785
C. Homog.	2,7352	3,3051	3,4767	4,8815	12,447,7111	3,690,2188	7,119,0005	2,473,3564	99,2484	186,3093	126,6278	7,6001	0,4385	3,1417
C. Curtosis	3,7352	4,3051	4,4767	5,8815	12,448,7111	3,691,2188	7,120,0005	2,474,3564	100,2484	187,3093	127,6278	8,6001	1,4385	4,1417

Fuente: Autor.

Con el análisis del cuadro anterior, se observa que el número máximo de fallas a clase disminuye drásticamente, y la asistencia a asesorías disminuye por 1.

La matriz de correlaciones generada para los datos se muestra a continuación.

Cuadro 6. Matriz de correlaciones 2.

Matriz de correlaciones (R)														
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas3	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
nota1	1,0000	0,4411	0,4223	0,6607	-0,0031	-0,0236	-0,0131	-0,0242	-0,0297	-0,0101	-0,0214	-0,0670	-0,3134	-0,3635
nota2	0,4411	1,0000	0,5503	0,7473	-0,0032	-0,0158	-0,0007	-0,0239	-0,0142	0,0061	-0,0013	-0,0403	0,1729	0,2138
nota3	0,4223	0,5503	1,0000	0,8768	-0,0066	-0,0276	-0,0031	-0,0289	0,0240	0,0579	0,0580	-0,0563	0,5734	0,6909
nota4	0,6607	0,7473	0,8768	1,0000	-0,0066	-0,0244	-0,0031	-0,0291	-0,0040	0,0310	0,0237	-0,0619	0,2947	0,3742
fallas1	-0,0031	-0,0032	-0,0066	-0,0066	1,0000	0,1287	0,0682	0,2052	-0,0021	-0,0016	-0,0023	0,0193	-0,0005	-0,0043
fallas2	-0,0236	-0,0158	-0,0276	-0,0244	0,1287	1,0000	0,2355	0,7082	-0,0030	0,0070	0,0044	0,0313	-0,0147	-0,0095
fallas3	-0,0131	-0,0007	-0,0031	-0,0031	0,0682	0,2355	1,0000	0,6187	-0,0018	-0,0023	-0,0027	0,0232	0,0078	0,0073
fallas4	-0,0242	-0,0239	-0,0289	-0,0291	0,2052	0,7082	0,6187	1,0000	-0,0038	0,0005	-0,0013	0,0399	-0,0078	-0,0104
nTalleres	-0,0297	-0,0142	0,0240	-0,0040	-0,0021	-0,0030	-0,0018	-0,0038	1,0000	0,1848	0,5926	-0,0128	0,0473	0,0484
nAsesorias	-0,0101	0,0061	0,0579	0,0310	-0,0016	0,0070	-0,0023	0,0005	0,1848	1,0000	0,9012	-0,0429	0,0627	0,0676
nTalleresM	-0,0214	-0,0013	0,0580	0,0237	-0,0023	0,0044	-0,0027	-0,0013	0,5926	0,9012	1,0000	-0,0408	0,0722	0,0767
edad	-0,0670	-0,0403	-0,0563	-0,0619	0,0193	0,0313	0,0232	0,0399	-0,0128	-0,0429	-0,0408	1,0000	0,0053	-0,0044
correlxnota	-0,3134	0,1729	0,5734	0,2947	-0,0005	-0,0147	0,0078	-0,0078	0,0473	0,0627	0,0722	0,0053	1,0000	0,8392
m	-0,3635	0,2138	0,6909	0,3742	-0,0043	-0,0095	0,0073	-0,0104	0,0484	0,0676	0,0767	-0,0044	0,8392	1,0000

Fuente: Autor.

Se observa que los estudiantes no fallan tanto a estas clases, y los que fallan, lo hacen más que todo en el segundo y tercer cohorte.

Sigue son haber correlación entre la edad y las demás variables, y la correlación entre nAsesorias y nTalleres con m aumenta en 0,04 aproximadamente.

Para hacer más específico el análisis descriptivo, ahora solo se tomarán los datos de los estudiantes que hayan asistido a talleres o asesorías de las asignaturas mencionadas anteriormente. El número de registros disminuye a 1.391.

A continuación se muestra el cuadro con el análisis descriptivo realizado.

Cuadro 7. Análisis descriptivo 3.

Descripción de los datos													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
Máximo	5,00	5,00	5,00	5,00	1,00	1,00	2,00	7,00	26,00	26,00	50,47	1,00	2,05
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	16,45	-1,00	-2,15
Rango	5,00	5,00	5,00	5,00	1,00	1,00	2,00	7,00	26,00	25,00	34,02	2,00	4,20
\bar{X}^2_{1x0}	3,18	3,20	3,44	3,34	0,00	0,00	0,00	0,65	1,55	2,20	22,29	0,17	0,13
Varianza	0,68	0,71	0,91	0,43	0,00	0,00	0,00	1,40	3,96	4,76	19,18	0,51	0,27
Desviación	0,83	0,84	0,95	0,65	0,03	0,04	0,06	1,18	1,99	2,18	4,38	0,71	0,52
C. Variab.	0,2602	0,2642	0,2769	0,1956	37,2961	26,3628	27,7909	1,8258	1,2832	0,9929	0,1965	4,1044	3,9593
C. Asimetrí	-0,6643	-0,5026	-1,0398	-0,8323	37,2157	26,2870	29,9408	2,4196	3,8967	3,6980	2,1785	-0,3773	-0,1200
C. Homog.	3,6558	3,1775	4,0890	5,0703	1.386,0043	691,5047	941,6525	8,4817	28,6231	25,6097	8,9004	0,6109	3,3778
C. Curtosis	4,6558	4,1775	5,0890	6,0703	1.387,0043	692,5047	942,6525	9,4817	29,6231	26,6097	9,9004	1,6109	4,3778

Fuente: Autor.

En el cuadro anterior se puede observar que las fallas siguen disminuyendo.

La matriz de correlaciones generada para los datos se muestra a continuación.

Cuadro 8. Matriz de correlaciones 3.

Matriz de correlaciones (R)													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
nota1	1,0000	0,3196	0,3288	0,6003	-0,0091	-0,0014	-0,0049	-0,0681	0,0264	-0,0128	-0,0603	-0,3712	-0,4962
nota2	0,3196	1,0000	0,3753	0,6904	-0,0317	-0,0201	-0,0268	-0,0867	-0,0005	-0,0474	-0,0525	0,1174	0,0898
nota3	0,3288	0,3753	1,0000	0,8290	-0,0462	-0,0375	-0,0444	-0,0022	0,1284	0,1159	-0,0920	0,5854	0,6568
nota4	0,6003	0,6904	0,8290	1,0000	-0,0427	-0,0285	-0,0371	-0,0739	0,0872	0,0395	-0,0932	0,2710	0,2829
fallas1	-0,0091	-0,0317	-0,0462	-0,0427	1,0000	0,7069	0,8944	-0,0147	-0,0074	-0,0147	0,0921	-0,0438	-0,0353
fallas2	-0,0014	-0,0201	-0,0375	-0,0285	0,7069	1,0000	0,9486	-0,0208	-0,0010	-0,0121	0,0859	-0,0623	-0,0334
fallas4	-0,0049	-0,0268	-0,0444	-0,0371	0,8944	0,9486	1,0000	-0,0197	-0,0039	-0,0143	0,0955	-0,0590	-0,0369
nTalleres	-0,0681	-0,0867	-0,0022	-0,0739	-0,0147	-0,0208	-0,0197	1,0000	-0,1261	0,4265	0,0505	0,0666	0,0523
nAsesorias	0,0264	-0,0005	0,1284	0,0872	-0,0074	-0,0010	-0,0039	-0,1261	1,0000	0,8434	-0,0382	0,1016	0,0969
nTalleresM	-0,0128	-0,0474	0,1159	0,0395	-0,0147	-0,0121	-0,0143	0,4265	0,8434	1,0000	-0,0075	0,1288	0,1167
edad	-0,0603	-0,0525	-0,0920	-0,0932	0,0921	0,0859	0,0955	0,0505	-0,0382	-0,0075	1,0000	-0,0021	-0,0364
correlxnota	-0,3712	0,1174	0,5854	0,2710	-0,0438	-0,0623	-0,0590	0,0666	0,1016	0,1288	-0,0021	1,0000	0,8345
m	-0,4962	0,0898	0,6568	0,2829	-0,0353	-0,0334	-0,0369	0,0523	0,0969	0,1167	-0,0364	0,8345	1,0000

Fuente: Autor.

En el análisis de correlaciones se puede observar que la relación entre nTalleres y nAsesorias con m sigue creciendo, con un aumento de 0,04 aproximadamente.

Se observa que la relación entre la edad y las fallas aumenta en 0,06 aproximadamente.

Se hace otra reducción del grupo para el análisis, tomando ahora solo los registros que tengan notas diferentes de cero en nota1, nota2 y nota3. El número de registros disminuye a 1.346.

A continuación se muestra el cuadro con el análisis descriptivo realizado.

Cuadro 9. Análisis descriptivo 4.

Descripción de los datos													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
Máximo	5,00	5,00	5,00	5,00	1,00	1,00	2,00	7,00	26,00	26,00	50,47	1,00	1,65
Mínimo	0,50	0,50	0,10	1,00	0,00	0,00	0,00	0,00	0,00	1,00	16,45	-1,00	-1,70
Rango	4,50	4,50	4,90	4,00	1,00	1,00	2,00	7,00	26,00	25,00	34,02	2,00	3,35
\bar{X}^2_{1x0}	3,23	3,24	3,51	3,39	0,00	0,00	0,00	0,65	1,56	2,21	22,24	0,19	0,14
Varianza	0,55	0,61	0,69	0,33	0,00	0,00	0,00	1,41	3,97	4,78	18,96	0,50	0,22
Desviación	0,74	0,78	0,83	0,58	0,03	0,04	0,06	1,19	1,99	2,19	4,35	0,71	0,47
C. Variab.	0,2290	0,2410	0,2366	0,1701	36,6879	25,9326	27,3374	1,8358	1,2731	0,9892	0,1958	3,8018	3,3578
C. Asimetrí	-0,0880	-0,0791	-0,4819	-0,0539	36,6061	25,8555	29,4498	2,4213	3,8839	3,6821	2,2152	-0,4064	-0,0989
C. Homog.	1,9357	1,9404	2,4558	2,5334	1.341,0045	669,0048	911,0527	8,4694	28,7408	25,6623	9,2052	0,6473	2,4074
C. Curtosis	2,9357	2,9404	3,4558	3,5334	1.342,0045	670,0048	912,0527	9,4694	29,7408	26,6623	10,2052	1,6473	3,4074

Fuente: Autor.

Se observa que dada la nueva restricción, los valores de nota1, nota2 y nota 3 son mayores que cero, el resto de variables mantiene los mismos extremos.

La matriz de correlaciones generada para los datos se muestra a continuación.

Cuadro 10. Matriz de correlaciones 4.

Matriz de correlaciones (R)													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresMasAsesorias	edad	correlxnota	m
nota1	1,0000	0,3097	0,2922	0,6283	-0,0123	-0,0044	-0,0083	-0,0718	0,0098	-0,0301	-0,0606	-0,4220	-0,5307
nota2	0,3097	1,0000	0,3022	0,6827	-0,0364	-0,0242	-0,0316	-0,1001	-0,0207	-0,0731	-0,0199	0,0740	0,0233
nota3	0,2922	0,3022	1,0000	0,7969	-0,0562	-0,0471	-0,0549	0,0066	0,1337	0,1253	-0,0675	0,5892	0,6555
nota4	0,6283	0,6827	0,7969	1,0000	-0,0515	-0,0360	-0,0458	-0,0783	0,0748	0,0257	-0,0708	0,2267	0,2101
fallas1	-0,0123	-0,0364	-0,0562	-0,0515	1,0000	0,7068	0,8944	-0,0149	-0,0077	-0,0151	0,0944	-0,0453	-0,0401
fallas2	-0,0044	-0,0242	-0,0471	-0,0360	0,7068	1,0000	0,9486	-0,0210	-0,0013	-0,0125	0,0883	-0,0645	-0,0382
fallas4	-0,0083	-0,0316	-0,0549	-0,0458	0,8944	0,9486	1,0000	-0,0199	-0,0043	-0,0147	0,0981	-0,0610	-0,0421
nTalleres	-0,0718	-0,1001	0,0066	-0,0783	-0,0149	-0,0210	-0,0199	1,0000	-0,1253	0,4284	0,0585	0,0720	0,0626
nAsesorias	0,0098	-0,0207	0,1337	0,0748	-0,0077	-0,0013	-0,0043	-0,1253	1,0000	0,8428	-0,0327	0,1025	0,1107
nTalleresM	-0,0301	-0,0731	0,1253	0,0257	-0,0151	-0,0125	-0,0147	0,4284	0,8428	1,0000	0,0020	0,1324	0,1348
edad	-0,0606	-0,0199	-0,0675	-0,0708	0,0944	0,0883	0,0981	0,0585	-0,0327	0,0020	1,0000	0,0116	-0,0120
correlxnota	-0,4220	0,0740	0,5892	0,2267	-0,0453	-0,0645	-0,0610	0,0720	0,1025	0,1324	0,0116	1,0000	0,8554
m	-0,5307	0,0233	0,6555	0,2101	-0,0401	-0,0382	-0,0421	0,0626	0,1107	0,1348	-0,0120	0,8554	1,0000

Fuente: Autor.

En la matriz de correlaciones se observa que sigue aumentando la correlación entre nTalleres y nAsesorías con m, esta vez con un aumento de 0,02 aproximadamente.

Y finalmente, se hará el mismo análisis descriptivo pero solo para las materias 'Cálculo I', 'Cálculo II', 'Cálculo III', 'Matemáticas', 'Matemáticas II'. Reduciendo los registros a 909.

A continuación se muestra el cuadro con el análisis descriptivo realizado.

Cuadro 11. Análisis descriptivo 5.

Descripción de los datos													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresM	edad	correlxnota	m
Máximo	5,00	5,00	5,00	5,00	1,00	1,00	2,00	7,00	26,00	26,00	50,47	1,00	1,55
Mínimo	0,50	0,50	0,10	1,40	0,00	0,00	0,00	0,00	0,00	1,00	16,45	-1,00	-1,70
Rango	4,50	4,50	4,90	3,60	1,00	1,00	2,00	7,00	26,00	25,00	34,02	2,00	3,25
\bar{X}	3,18	3,22	3,48	3,34	0,00	0,00	0,00	0,94	1,51	2,45	22,09	0,22	0,15
Varianza	0,47	0,52	0,70	0,32	0,00	0,00	0,01	1,80	4,74	5,82	19,65	0,51	0,21
Desviación	0,68	0,72	0,84	0,57	0,03	0,05	0,07	1,34	2,18	2,41	4,43	0,71	0,46
C. Variab.	0,2155	0,2230	0,2401	0,1692	30,1496	21,3073	22,4623	1,4318	1,4416	0,9858	0,2007	3,2399	3,0318
C. Asimetría	0,0191	-0,0552	-0,5105	0,0328	30,0502	21,2135	24,1686	1,8673	3,8517	3,4558	2,2200	-0,4794	-0,2358
C. Homog.	2,0029	2,2624	2,6527	2,4490	904,0066	450,5071	613,8949	5,6377	27,3694	22,7622	9,3218	0,6939	2,6927
C. Curtosis	3,0029	3,2624	3,6527	3,4490	905,0066	451,5071	614,8949	6,6377	28,3694	23,7622	10,3218	1,6939	3,6927

Fuente: Autor.

No se observan cambios significativos en el análisis.

La matriz de correlaciones generada para los datos se muestra a continuación.

Cuadro 12. Matriz de correlaciones 5.

Matriz de correlaciones (R)													
Variable	nota1	nota2	nota3	nota4	fallas1	fallas2	fallas4	nTalleres	nAsesorias	nTalleresMasAsesorias	edad	correlxnota	m
nota1	1,0000	0,3938	0,2709	0,6313	-0,0134	-0,0018	-0,0072	-0,0354	0,0008	-0,0190	-0,1116	-0,4095	-0,4948
nota2	0,3938	1,0000	0,3583	0,7113	-0,0472	-0,0308	-0,0406	-0,1172	-0,0325	-0,0945	-0,0435	0,0787	0,0324
nota3	0,2709	0,3583	1,0000	0,8129	-0,0668	-0,0552	-0,0648	0,0415	0,1278	0,1384	-0,0751	0,6054	0,7025
nota4	0,6313	0,7113	0,8129	1,0000	-0,0611	-0,0408	-0,0532	-0,0392	0,0617	0,0339	-0,0980	0,2580	0,2672
fallas1	-0,0134	-0,0472	-0,0668	-0,0611	1,0000	0,7067	0,8943	-0,0232	-0,0078	-0,0199	0,1141	-0,0563	-0,0504
fallas2	-0,0018	-0,0308	-0,0552	-0,0408	0,7067	1,0000	0,9486	-0,0328	-0,0002	-0,0185	0,1072	-0,0801	-0,0485
fallas4	-0,0072	-0,0406	-0,0648	-0,0532	0,8943	0,9486	1,0000	-0,0311	-0,0036	-0,0206	0,1189	-0,0759	-0,0532
nTalleres	-0,0354	-0,1172	0,0415	-0,0392	-0,0232	-0,0328	-0,0311	1,0000	-0,1232	0,4450	0,1077	0,0607	0,0636
nAsesorias	0,0008	-0,0325	0,1278	0,0617	-0,0078	-0,0002	-0,0036	-0,1232	1,0000	0,8339	-0,0440	0,1093	0,1148
nTalleresM	-0,0190	-0,0945	0,1384	0,0339	-0,0199	-0,0185	-0,0206	0,4450	0,8339	1,0000	0,0203	0,1323	0,1390
edad	-0,1116	-0,0435	-0,0751	-0,0980	0,1141	0,1072	0,1189	0,1077	-0,0440	0,0203	1,0000	0,0310	0,0147
correlxnota	-0,4095	0,0787	0,6054	0,2580	-0,0563	-0,0801	-0,0759	0,0607	0,1093	0,1323	0,0310	1,0000	0,8493
m	-0,4948	0,0324	0,7025	0,2672	-0,0504	-0,0485	-0,0532	0,0636	0,1148	0,1390	0,0147	0,8493	1,0000

Fuente: Autor.

No se observan cambios significativos en el análisis. Aunque el aumento de la correlación entre las fallas y la edad aumenta en 0,02.

10.2 ANÁLISIS EMPRESAS DONDE TRABAJAN LOS ESTUDIANTES

Se hace un análisis de la tabla T_Matricula2 mostrando las diferentes empresas en las que trabajan los estudiantes matriculados y se observa que hay varias entidades o empresas que son la misma, pero aparecen con los nombres escritos de maneras diferentes. En la siguiente imagen se muestra un ejemplo con la entidad Policía Nacional.

Cuadro 13. Nombres con que se registra la entidad Policía Nacional.

datp_NombreEmpresa		
Policía Nacional	POLICIA NACIONAL COLOMBIA	MINISTERIO DE DEFENZA POLICIA NACIONAL
Policía Nacional	Policia Nacional Área de Aviacion	LA POLICIA
POLICIA NACIONBAL	Policia Nacional- Aeropuerto	INSPECCIÓN GENERAL DE LA POLICIA NACIONAL
Policia Nacional-DIJIN	POLICIA NACIONAL - METROPOLITANA DE TUNJA	HOSPITAL DE LA POLICIA
POLICIA NACIONAL, AREA DE SANIDAD	POLICIA NACIONAL - DEPARTAMENTO DE POLICIA CESAR	FONDO ROTATORIO DE LA POLICÍA
POLICIA NACIONAL VIRTUAL	POLICIA NACIONAL - CENOP	FONDO ROTATORIO DE LA POLICIA
POLICIA NACIONAL SIJIN	POLICIA NACIONAL - AVIACION	ESCULA DE POLICIA NACIONAL
Policia Nacional FUCOT	POLICIA NACIONAL zambrano.1491@hotmail.com	DIRECCION DE SANIDAD POLICIA NACIONAL
POLICIA NACIONAL DIJIN	POLICIA NACIONAL JOANPEREZ03@HOTMAIL.COM	Dirección de Bienestar Social Policía Nacional
POLICIA NACIONAL DEANT	POLICIA NACIONAL joalbo-92@hotmail.com	DEPARTAMENTO POLICIA NACIONAL CHOCO
POLICIA NACIONAL DE MEDELLIN	Policia Nacional	comando de policia
Policia Nacional de los Colombianos	POLICIA NACIONAL	COLEGIO POLICIA NACIONAL
policia nacional de los colombianos	POLICIA NACIOANL	CLUB DE SUBOFICIALES DE LA POLICÍA NACIONAL
Policia Nacional de Colombia -	POLICIA NACINAL	CLINICA DE LA POLICIA
Policia Nacional de Colombia	POLICÍA NACIONAL	CAJA PROMOTORA DE VIVIENDA MILITAR Y DE POLICÍA
policia nacional de colombia	POLICIA NACIONAL	CAJA PROMOTORA DE VIVIENDA MILITAR Y DE POLICIA
Policia Nacional de Colombia	POLICIA	CAJA PROMOTORA DE VIVIENDA MILITAR Y DE POLICIA
POLICIA NACIONAL COLOMBIANA	OFICIALES DE LA POLICIA NACIONAL	

Fuente: Autor.

Para realizar un análisis de las empresas de donde provienen los estudiantes, es necesario estandarizar el nombre de la empresa o entidad por medio de una tabla, dado que este caso se presenta para la mayoría de los registros (aerolíneas, bancos, etc.).

En total aparecen registrados 26.608 empresas con distintos nombres entre los periodos 20161 y 20191.

11. OBSERVACIONES

Se hizo gestión para recibir la tabla del calendario académico para saber las asesorías o talleres a los que asisten los estudiantes antes del parcial de cada corte y notar mejor la relación existente entre asesorías y notas parciales.

También se hizo gestión para tener acceso directo a algunas tablas o bodegas de datos y utilizar el servicios de la universidad para el procesamiento previo de la información. Pero no se logró tener el acceso directo a las tablas de la base de datos, por lo que el trabajo se realizó solo con la información recibida del área de permanencia.

El análisis de las empresas desde las que más provienen los estudiantes matriculados no se puede hacer dada la baja calidad de los datos en el campo datp_NombreEmpresa en la tabla T_Matricula2.

12. RECOMENDACIONES

Dado que se observa una correlación entre la asistencia a asesorías extra clase y talleres con la tendencia creciente de la nota, se recomienda explorar qué métodos están haciendo que las asesorías y talleres sean efectivos para seguirlos usando, y buscar nuevos métodos que permitan hacer más eficientes estos espacios para los estudiantes.

También se recomienda la creación de una bodega de datos con los datos de los interesados en estudiar y los estudiantes matriculados del Politécnico Grancolombiano para realizar un análisis del perfil de ingreso a la Universidad con el fin de establecer convenios con diferentes entidades para otorgar algunos beneficios a los nuevos estudiantes provenientes de dichas entidades.

La bodega de datos no serviría solo para realizar análisis de minería, sino que también serviría para crear algunos reportes que permitan ver la información de la universidad de manera automática sin gastar dinero en procesamiento de datos de forma manual.

Con la estandarización de los nombres de las empresas de las que provienen los estudiantes matriculados, se podría luego analizar las empresas en las que trabaja la mayoría de los estudiantes para realizar algunos convenios nuevos o fortalecer los convenios existentes.