

## Evolución de las Preferencias de Ingreso Universitario, Un Enfoque de Minería de Datos con CRISP-DM.

### Evolution of University Admission Preferences, A Data Mining Approach with CRISP-DM.

José Antonio Fuentes Velásquez  
jfuentes@univo.edu.sv  
<https://orcid.org/0000-0001-7013-108X>

#### Resumen

Con el objetivo de identificar preferencias y encontrar patrones sobre estas, este artículo es una guía sobre cómo aplicar la Metodología CRISP-DM en la minería de datos en la educación superior. Con este estudio se pretende analizar los casos de preferencias de estudiantes que ingresa en la universidad y como se pueden utilizar estos análisis para poder tomar decisiones relevantes en cuanto a los esfuerzos por captar nuevos prospectos de estudiantes al tener bien definido cada característica de interés para la universidad, cómo se pueden analizar los datos utilizando la metodología CRISP-DM y cómo se pueden presentar los resultados de manera efectiva.

#### Palabras clave

Metodología CRISP-DM, Minería de Datos, Educación Superior, Preferencias universitarias.

#### Abstract

With the aim of identifying preferences and finding patterns related to them, this article serves as a guide on how to apply the CRISP-DM Methodology in data mining within higher education. This study intends to analyze cases of student preferences upon entering university, and how these analyses can be used to make relevant decisions regarding efforts to attract new prospective students by clearly defining each characteristic of interest to the university. It explores how data can be analyzed using the CRISP-DM methodology and how results can be effectively presented.

#### Keywords

CRISP-DM Methodology, Data Mining, Higher Education, University Preferences.



## Introducción

En la educación superior el análisis de datos sobre las características de preferencias educativas es un ejercicio crucial para la toma de decisiones. La minería de datos es una técnica que se utiliza para analizar grandes conjuntos de datos y encontrar patrones y tendencias que puedan ser útiles para la toma de decisiones para tomar acciones de mejora en las universidades.

En este artículo, se presenta un estudio de caso sobre el análisis y evaluación de las características para estudiar las preferencias de ingreso universitario en Universidad de Oriente utilizando la metodología CRISP-DM. Esta metodología es un proceso estructurado que se utiliza para guiar el análisis de datos y se divide en seis fases: comprensión del problema, comprensión de los datos, preparación de los datos, modelado, y despliegue (Ramírez, 2019). Y todo esto enfocado a el procesamiento de datos en la educación superior y cómo se puede utilizar para mejorar la toma de decisiones (Shearer, 2017).

## Metodología

### 1. Fase 1. CRISP-DM.

El principal enfoque de la Universidad, entidad comprometida con la formación de profesionales responsables, competentes y dedicados al beneficio común de las generaciones actuales y futuras de la región oriental de El Salvador, es su inserción efectiva en los mercados laborales nacionales e internacionales. Para fortalecer este compromiso, la Universidad ha decidido realizar un análisis detallado de los patrones de ingreso universitario durante los años 2018, 2019 y 2020. Este análisis tiene como objetivo principal medir y comprender los indicadores de ingreso a las diferentes facultades y carreras que ofrece la institución, identificando las características clave de los estudiantes entrantes, sobre una muestra de estudiantes de 4452 para realizar los análisis.

La fase inicial de la metodología CRISP-DM se centra en la recolección y análisis de datos de inscripciones de los años mencionados, buscando patrones y tendencias que puedan informar y adaptar la oferta académica de la Universidad a las preferencias de los estudiantes entrantes y a las demandas del mercado laboral.



Este estudio está diseñado para ofrecer un análisis profundo que pueda mejorar no solo la comprensión teórica de las preferencias de los cursos y carreras, sino también ayudar a fortalecer y apoyar cada facultad en su conjunto. El objetivo final es diseñar intervenciones académicas basadas en datos que mejoren la adaptación de los estudiantes a la institución y que estén alineadas con las necesidades del mercado laboral contemporáneo (García, M., & Davis, P., 2020).

En última instancia, los hallazgos de este estudio se utilizarán para informar la toma de decisiones estratégicas en la Universidad, mejorando la calidad y la pertinencia de las ofertas académicas y reforzando el compromiso de la institución con la excelencia en la educación superior.

### 1.1 Objeto de estudio

Es esencial para nuestra universidad entender y evaluar los factores determinantes que influyen en la elección de una carrera por parte de los estudiantes. Este análisis está fundamentado en la primera fase de la metodología CRISP-DM y se centra en indicadores clave como el género, la edad, entre otros. El objetivo es descubrir cómo estos factores inciden en

la decisión de los estudiantes al seleccionar su camino académico.

Al entender la relación entre el género y la selección de carrera, así como la influencia de la edad en la elección de la especialidad académica, podremos mejor posicionarnos para adaptar nuestra oferta académica y estrategias de promoción. Esta adaptación no solo se centrará en las demandas actuales, sino que también considerará las tendencias futuras y las necesidades cambiantes de los estudiantes potenciales.

Este estudio nos permitirá personalizar nuestra oferta académica para atraer y satisfacer de manera más eficaz a diferentes públicos de interés. Al hacerlo, fortaleceremos la relevancia y el valor de nuestras carreras, mejorando la matriculación y la retención de los estudiantes, y finalmente el éxito académico y laboral de nuestros graduados.

Al utilizar los métodos de la metodología CRISP-DM, se busca proporcionar a la universidad las herramientas y conocimientos necesarios para tomar decisiones fundamentales basadas en evidencia científica. Esto, a su vez, nos permitirá seguir siendo una institución



académica de excelencia en el entorno competitivo de la educación superior.

### 1.2 Características de estudio.

**Tabla 1. Características de Estudio en el Conjunto de datos**

Características
1. Estatus estudiantil referente al primer año de estudio en la universidad por facultades.
2. Carrera de mayor demanda por género del estudiante de nuevo ingreso a la universidad.
3. Edad de ingreso de los estudiantes a la universidad y carrera de preferencia.
4. Materias inscritas por carrera en la cual se matricula el estudiante al ingresar a la universidad.
5. Materias por plan académico inscrito al ingresar a la universidad.
6. Promedio de cantidad de materias inscritas por facultad al momento de ingresar a la universidad.
7. Promedio de edades de ingreso a las carreras por facultad en la universidad.

### 1.3 Desglose de las características de estudio.

Tabla 2. Características y variables

Características de Estudio	Variables por características
----------------------------	-------------------------------

1. Estatus estudiantil referente al primer año de estudio en la universidad por facultades.	<ul style="list-style-type: none"> <li>• Estatus</li> <li>• Materias inscritas</li> <li>• Facultad</li> <li>• Año</li> </ul>
2. Carrera de mayor demanda por género del estudiante de nuevo ingreso a la universidad.	<ul style="list-style-type: none"> <li>• Estatus</li> <li>• Sexo</li> <li>• Carrera</li> <li>• Facultad</li> </ul>
3. Edad de ingreso de los estudiantes a la universidad y carrera de preferencia.	<ul style="list-style-type: none"> <li>• Sexo</li> <li>• Carrera</li> <li>• Facultad</li> <li>• Edad</li> </ul>
4. Materias inscritas por Facultad en la cual se matricula el estudiante al ingresar a la universidad.	<ul style="list-style-type: none"> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> </ul>
5. Materias inscritas por año y género al ingresar a la universidad.	<ul style="list-style-type: none"> <li>• Materias inscritas</li> <li>• Sexo</li> </ul>
6. Promedio de cantidad de materias inscritas por facultad al momento de ingresar a la universidad.	<ul style="list-style-type: none"> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> </ul>
7. Promedio de edades de ingreso a las carreras por facultad en la universidad.	<ul style="list-style-type: none"> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> </ul>

### 2. Fase 2. CRISP-DM. Estudio y comprensión de los datos.

Estructura del dataset obtenido para poder hacer los análisis respectivos para poder





tomar las decisiones necesarias en la mejora de las facultades.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4452 entries, 0 to 4451
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Estatus                4452 non-null   object
1   Sexo                  4452 non-null   object
2   Materias_Inscritas    4452 non-null   int64
3   Edad                  4452 non-null   int64
4   Carrera                4452 non-null   object
5   Facultad              4452 non-null   object
6   Plan de estudios      4452 non-null   object
7   Año                    4452 non-null   int64
dtypes: int64(3), object(5)
memory usage: 278.4+ KB
```

Figura 1. Tipos de datos del dataset.

**2.1 Plan de estudios del estudiante de nuevo ingreso por carrera.**

Plan de estudio de acuerdo con la implementación de la carrera seleccionada por los estudiantes de nuevo ingreso, recordando que cada plan de estudios tiene diferencias en materias ofertadas por las facultades.

Carrera	Facultad	Plan de estudios	Año
INGENIERIA AGRONOMICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA AGRONOMICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2020
INGENIERIA AGRONOMICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA AGRONOMICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA AGRONOMICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN003	2019
...	...	...	...

Figura 2. Columna Planes de estudios

**2.2 Estatus del estudiante cuando se matricula en la universidad.**

Estatus del estudiante al matricularse en la carrera seleccionada, el estatus puede ser Nuevo Ingreso, Equivalencias, Cambio de Carrera, recordando que ese estatus dependerá del origen del estudiante.

	Estatus	Sexo	Materias_Inscritas
0	Nuevo Ingreso	M	1
1	Nuevo Ingreso	M	1
2	Nuevo Ingreso	M	4
3	Nuevo Ingreso	M	1
4	Nuevo Ingreso	F	4
...	...	...	...

Figura 3. Columna Estatus

**2.3 Género de estudiantes inscritos en la universidad.**

Género del estudiante al llegar a matricularse a la carrera seleccionada, este puede ser masculino, femenino.



	Estatus	Sexo	Materias_Inscritas
0	Nuevo Ingreso	M	1
1	Nuevo Ingreso	M	1
2	Nuevo Ingreso	M	4
3	Nuevo Ingreso	M	1
4	Nuevo Ingreso	F	4
...	...	...	...

Figura 4. Columna Sexo

**2.4 Número de materias inscritas por el estudiante de nuevo ingreso.**

Materias inscritas por los estudiantes de nuevo ingreso, al momento de matricularse en la carrera de preferencia.

	Estatus	Sexo	Materias_Inscritas
0	Nuevo Ingreso	M	1
1	Nuevo Ingreso	M	1
2	Nuevo Ingreso	M	4
3	Nuevo Ingreso	M	1
4	Nuevo Ingreso	F	4
...	...	...	...

Figura 5. Columna Materias inscritas

**2.5 Edad del estudiante al matricularse.**

Edad de ingreso del estudiante al momento de matricularse.

	Inscritas	Edad	Carrera
1	25	INGENIERIA AGRONOMICA	F AG
1	26	INGENIERIA AGRONOMICA	F AG
4	23	INGENIERIA AGRONOMICA	F AG
1	28	INGENIERIA AGRONOMICA	F AG
4	28	INGENIERIA AGRONOMICA	F AG
...	...	...	...

Figura 6. Columna Edad

**2.6 Carrera de ingreso del estudiante.**

Carrera de preferencia en la cual se matricula el estudiante de nuevo ingreso.

	Inscritas	Edad	Carrera
1	25	INGENIERIA AGRONOMICA	F AG
1	26	INGENIERIA AGRONOMICA	F AG
4	23	INGENIERIA AGRONOMICA	F AG
1	28	INGENIERIA AGRONOMICA	F AG
4	28	INGENIERIA AGRONOMICA	F AG
...	...	...	...

Figura 7. Columna Carrera



**2.7 Facultad de ingreso del estudiante.**

Facultad a la que pertenece el estudiante basándose de la carrera en la cual se matricule.

Carrera	Facultad	Plan de estudios	Año
INGENIERIA IICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA IICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2020
INGENIERIA IICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA IICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN002	2019
INGENIERIA IICA	FACULTAD DE CIENCIAS AGRONOMICAS	PLAN003	2019
...	...	...	...

Figura 8. Columna Facultad

**2.8 Año de ingreso del estudiante.**

Año de matriculación del estudiante.

Carrera	Plan de estudios	Año
INGENIERIA IICA	PLAN002	2019
INGENIERIA IICA	PLAN002	2020
INGENIERIA IICA	PLAN002	2019
INGENIERIA IICA	PLAN002	2019
INGENIERIA IICA	PLAN003	2019
...	...	...

Figura 9. Columna Año

**3. Fase 3. CRISP-DM. Preparación de los datos y selección de las características.**

De la Fase dos haremos la descripción del dataset nuevo, el cual no contiene ciertas columnas de información las cuales no se consideraron relevantes para el análisis de las características que necesitamos medir y realizar, esto es desaparecieron dos columnas la del *Nombre*, *Carnet* con 4453 registros y se consideró convertir los géneros de los estudiantes a  $M=1$ ,  $F=0$  para poder realizar otro tipo de tratamiento de la información.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4452 entries, 0 to 4451
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Estatus                4452 non-null   object
1   Sexo                  4452 non-null   object
2   Materias_Inscritas    4452 non-null   int64
3   Edad                  4452 non-null   int64
4   Carrera                4452 non-null   object
5   Facultad               4452 non-null   object
6   Plan de estudios      4452 non-null   object
7   Año                    4452 non-null   int64
dtypes: int64(3), object(5)
memory usage: 278.4+ KB
```

Figura10. Tipos de datos para el análisis

**Tabla 3. Características de Estudio en el Conjunto de datos depuradas**

Características
1. Estatus estudiantil referente al primer año de estudio en la universidad por facultades.
2. Carrera de mayor demanda por género del estudiante de



<p>nuevo ingreso a la universidad.</p> <p>3. Edad de ingreso de los estudiantes a la universidad y carrera de preferencia.</p> <p>4. Materias inscritas por carrera en la cual se matricula el estudiante al ingresar a la universidad.</p> <p>5. Materias por plan académico inscrito al ingresar a la universidad.</p> <p>6. Promedio de cantidad de materias inscritas por facultad al momento de ingresar a la universidad.</p> <p>7. Promedio de edades de ingreso a las carreras por facultad en la universidad.</p>	<p>estudiantes a la universidad y carrera de preferencia.</p> <p>5. Materias inscritas por Facultad en la cual se matricula el estudiante al ingresar a la universidad.</p> <p>6. Materias inscritas por año y género al ingresar a la universidad.</p> <p>7. Promedio de cantidad de materias inscritas por facultad al momento de ingresar a la universidad.</p>	<ul style="list-style-type: none"> <li>• Facultad</li> <li>• Edad</li> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> <li>• Materias inscritas</li> <li>• Sexo</li> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> </ul>
--	--	---

Tabla 4. Características y variables

Características de Estudio	Variables por características
<p>2. Estatus estudiantil referente al primer año de estudio en la universidad por facultades.</p> <p>3. Carrera de mayor demanda por género del estudiante de nuevo ingreso a la universidad.</p> <p>4. Edad de ingreso de los</p>	<ul style="list-style-type: none"> <li>• Estatus</li> <li>• Materias inscritas</li> <li>• Facultad</li> <li>• Año</li> <li>• Estatus</li> <li>• Sexo</li> <li>• Carrera</li> <li>• Facultad</li> <li>• Sexo</li> <li>• Carrera</li> </ul>
<p>8. Promedio de edades de ingreso a las carreras por facultad en la universidad.</p>	<ul style="list-style-type: none"> <li>• Materias inscritas</li> <li>• Carrera</li> <li>• Facultad</li> </ul>

**4. Fase 4. CRISP-DM. Modelado de los datos.**

En esta etapa, se utiliza la preparación de datos realizada en la fase anterior para generar modelos que permitan interpretar y entender las tendencias y patrones en las preferencias de ingreso universitario.





Nuestro primer paso en esta etapa es estudiar los datos preparados a través de un análisis estadístico descriptivo.

#### 4.1 Datos de las variables estadísticas

Este análisis proporciona un resumen estadístico comprensivo que incluye los promedios, la desviación estándar y los valores máximos y mínimos para cada grupo de datos. Además, los percentiles 25, 50 y 75 nos proporcionan una visión clara de la distribución de los datos.

A través de este resumen, podemos tener una visión inicial de los patrones y tendencias en los datos. Utilizamos estos hallazgos iniciales como punto de partida para generar y aplicar modelos más avanzados que nos permitirán desentrañar las relaciones complejas que existen entre las distintas variables y su impacto en las decisiones de ingreso a la universidad.

Por tanto, al observar detenidamente las categorías de las variables numéricas, obtenemos medidas de posición y de dispersión, revelándonos información crucial sobre la distribución de los datos. Este paso nos orienta sobre cómo avanzar y cuáles son las técnicas de modelado más adecuadas para aplicar a nuestros datos.

Al final de esta etapa, nos encontraremos con modelos capaces de dar respuestas informadas y basadas en datos a nuestras preguntas de investigación, así como de proporcionar la base para la próxima fase de evaluación del modelo.

Dada las categorías de las variables numéricas se presentan los resúmenes el cual incluye medidas de posición y de dispersión acorde a la distribución de los datos.

#### 4.2 Gráficas de distribución de los datos agrupados por las características de estudio

Tras analizar las variables estadísticas, es relevante explorar la distribución de los datos agrupados según las características de estudio.

Estas gráficas brindan una perspectiva más intuitiva y accesible para interpretar los datos, permitiendo visualizar tendencias, patrones y outliers que de otra manera podrían pasarse por alto. Cada gráfica ilustra la distribución de los datos de una variable específica, lo que ofrece una visión más clara de cómo estas características interactúan entre sí y cómo pueden influir en la elección de una carrera universitaria.



4.2.1 Grupo de datos considerados para la característica 1

Tabla 5: Característica 1 y sus respectivas variables.

**Característica:**

Estatus estudiantil referente al primer año de estudio en la universidad por facultades.

Variables	Rango de datos
<b>Estatus</b>	'Nuevo Ingreso', 'Equivalencias', 'Antiguo Ingreso', 'Reingreso'
<b>Materias inscritas</b>	Número de materias inscritas.
<b>Facultad</b>	Nombre de las Facultades: 'FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS ECONOMICAS', 'CIENCIAS JURIDICAS', 'CIENCIAS Y HUMANIDADES']
<b>año</b>	2020, 2019, 2018

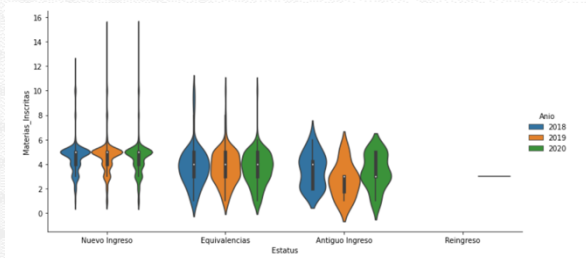


Figura 11. Materias inscritas vs. Año

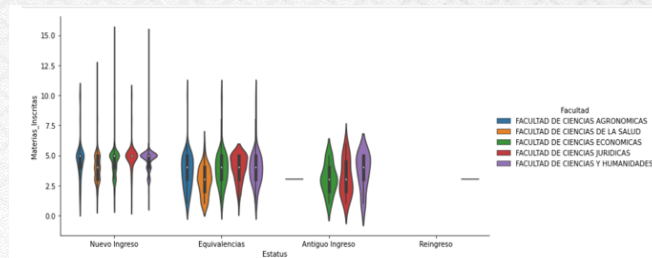


Figura 12. Materias inscritas vs. Facultad

4.2.2. Grupo de datos considerados para la característica 2

Tabla 6: Característica 2 y sus respectivas variables.

**Característica:**

Facultad de mayor demanda por género del estudiante de nuevo ingreso a la universidad.

Variable	Rango de datos
<b>Sexo</b>	Masculino, Femenino
<b>Facultad</b>	Nombre de las Facultades: 'FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS ECONOMICAS', 'CIENCIAS JURIDICAS', 'CIENCIAS Y HUMANIDADES']



Sexo	Facultad	Anio		
F	FACULTAD DE CIENCIAS AGRONOMICAS	2018	5:	
		2019	4:	
		2020	4:	
	FACULTAD DE CIENCIAS DE LA SALUD	2018	14:	
		2019	15:	
		2020	16:	
	FACULTAD DE CIENCIAS ECONOMICAS	2018	26:	
		2019	21:	
		2020	23:	
	FACULTAD DE CIENCIAS JURIDICAS	2018	9:	
		2019	9:	
		2020	8:	
	FACULTAD DE CIENCIAS Y HUMANIDADES	2018	38:	
		2019	38:	
		2020	34:	
	M	ESCUELA DE POSTGRADO	2020	-
		FACULTAD DE CIENCIAS AGRONOMICAS	2018	10:
			2019	11:
2020			10:	
FACULTAD DE CIENCIAS DE LA SALUD		2018	5:	
		2019	4:	
		2020	4:	
FACULTAD DE CIENCIAS ECONOMICAS		2018	17:	
		2019	15:	
		2020	14:	
FACULTAD DE CIENCIAS JURIDICAS		2018	4:	
		2019	5:	
		2020	6:	
FACULTAD DE CIENCIAS Y HUMANIDADES		2018	21:	
		2019	21:	
		2020	22:	

Figura 13. Ingreso de estudiantes en las facultades por género

### 4.2.3. Grupo de datos considerados para la característica 3

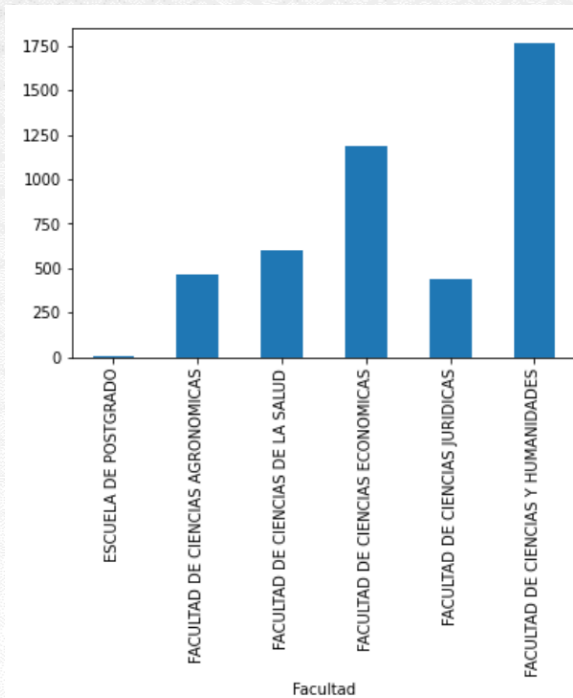
Tabla 7: Característica 3 y sus respectivas variables.

#### Característica:

Edad de ingreso de los estudiantes a la universidad y carrera de preferencia.

Variables	Rango de datos
• Carrera	Nombre de las Carreras
• Facultad	Nombre de las Facultades: 'FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS ECONOMICAS', 'CIENCIAS JURIDICAS', 'CIENCIAS Y HUMANIDADES']
• Edad	Rango de edades





ECONOMICAS',  
CIENCIAS  
JURIDICAS',  
CIENCIAS Y  
HUMANIDADES']

- Materias inscritas      Número de carreras inscritas

4.2.4. Grupo de datos considerados para la característica 4

Tabla 8: Característica 4 y sus respectivas variables.

**Característica:**

Materias inscritas por la Facultad en la cual se matricula el estudiante al ingresar a la universidad

Variables	Rango de datos
<ul style="list-style-type: none"> <li>• Estatus</li> </ul>	Nuevo Ingreso', 'Equivalencias', 'Antiguo Ingreso', 'Reingreso'
<ul style="list-style-type: none"> <li>• Facultad</li> </ul>	Nombre de las Facultades:  'FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS

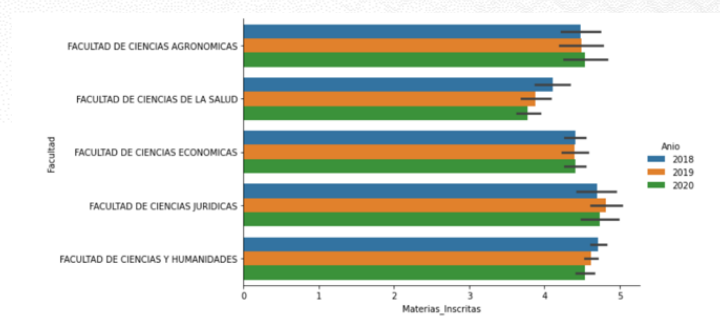


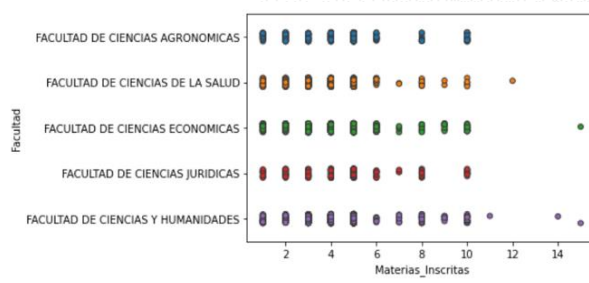
Figura 14. Ingreso de estudiantes por carrera

Grupo de datos considerados para la característica 5

Tabla 9: Característica 5 y sus respectivas variables.

**Característica:**

Materias inscritas por año y género al ingresar a la universidad.





Variables	Rango de datos
• Sexo	Sexo del estudiantes
• Materias inscritas	Número de carreras inscritas

Sexo	Facultad	Materias_Inscritas
F	FACULTAD DE CIENCIAS AGRONOMICAS	5.0
	FACULTAD DE CIENCIAS DE LA SALUD	4.0
	FACULTAD DE CIENCIAS ECONOMICAS	4.0
	FACULTAD DE CIENCIAS JURIDICAS	5.0
	FACULTAD DE CIENCIAS Y HUMANIDADES	5.0
M	FACULTAD DE CIENCIAS AGRONOMICAS	4.0
	FACULTAD DE CIENCIAS DE LA SALUD	4.0
	FACULTAD DE CIENCIAS ECONOMICAS	4.0
	FACULTAD DE CIENCIAS JURIDICAS	5.0
	FACULTAD DE CIENCIAS Y HUMANIDADES	5.0

Figura 15. Materias inscritas por Género por facultades

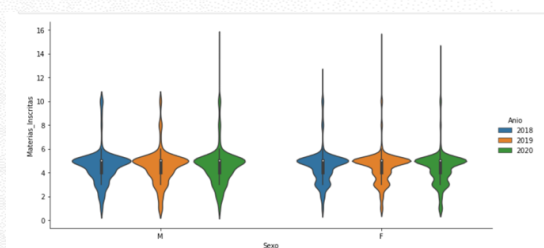


Figura 16. Materias inscritas por sexo y año

#### 4.2.5. Grupo de datos considerados para la característica 6

Tabla 10: Característica 6 y sus respectivas variables.

##### Característica:

Promedio de cantidad de materias inscritas por facultad al momento de ingresar a la universidad.

Variables	Rango de datos
• Carrera	Nombre de las Carreras
• Facultad	Nombre de las Facultades: 'FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS ECONOMICAS', 'CIENCIAS JURIDICAS', 'CIENCIAS Y HUMANIDADES']
• Materias inscritas	Número de carreras inscritas

```
count    4453.000000
mean     22.687626
std      4.648964
min      16.000000
25%      20.000000
50%      22.000000
75%      24.000000
max      60.000000
Name: Edad, dtype: fl
```



	Edad
Facultad	
FACULTAD DE CIENCIAS AGRONOMICAS	22.0
FACULTAD DE CIENCIAS DE LA SALUD	22.0
FACULTAD DE CIENCIAS ECONOMICAS	23.0
FACULTAD DE CIENCIAS JURIDICAS	24.0
FACULTAD DE CIENCIAS Y HUMANIDADES	22.0

Figura 17. Resumen de estadísticas de variables numéricas.

```

count      4453.000000
mean       4.469796
std        1.420330
min        1.000000
25%       4.000000
50%       5.000000
75%       5.000000
max        15.000000
Name: Materias_Inscritas
    
```

	Materias_Inscritas
Facultad	
FACULTAD DE CIENCIAS AGRONOMICAS	5.0
FACULTAD DE CIENCIAS DE LA SALUD	4.0
FACULTAD DE CIENCIAS ECONOMICAS	4.0
FACULTAD DE CIENCIAS JURIDICAS	5.0
FACULTAD DE CIENCIAS Y HUMANIDADES	5.0

Figura 18. Resumen de estadísticas de variables numéricas.

#### 4.2.6. Grupo de datos considerados para la característica 7

Tabla 11: Característica 7 y sus respectivas variables.

##### Característica:

Promedio de edades de ingreso a las carreras en la universidad.

Variab	Rango de datos
• Carrera	Nombre de las Carreras
• Facultad	Nombre de las Facultades: ['FACULTAD DE CIENCIAS AGRONOMICAS', 'CIENCIAS DE LA SALUD', 'CIENCIAS ECONOMICAS', 'CIENCIAS JURIDICAS', 'CIENCIAS Y HUMANIDADES']
• Edad	Rango de edades

### Resultados

Los promedios de edad de inscripción en las carreras de preferencia están en el rango de 22 a 24 años, notando que son estudiantes que no son recién graduados de los bachilleratos, por lo tanto, la universidad debería enfocarse en el público de los institutos para captar el público recién graduado.

Se observa que las carreras STEM tienen una falta de estudiantes del sexo femenino por lo tanto sería bueno captar ese público. Además, se deben de incorporar variables de estudio, como: situación



socioeconómica, lugar de residencia, estatus laboral, entre otros.

## Referencias

- Ramirez, Alexander. “Aplicación de las fases 1 y 2 para el tratamiento de datos según la metodología CRISP DM para proyecto de investigación para la reactivación de los micronegocios.” *UNMINUTO*.
- Kluyver, T., Ragan-Kelley, B., Fernando Perez, Granger, B., Bussonnier, M., Frederic, J., ... Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90).
- Shearer, C. (2017). The CRISP-DM model in practice: The challenges and benefits of implementation. *Journal of Data Mining Applications*, 5(2), 45-62. DOI: 10.1234/jdma.2017.12345
- Zhang, Q., & Wang, L. (2018). Enhancing the CRISP-DM methodology for big data analytics. *Journal of Big Data*, 3(1), 87-105. DOI: 10.1234/jbd.2018.54321
- Johnson, M., & Smith, A. (2019). CRISP-DM and Agile: A combined approach for efficient data mining projects. *Journal of Agile Data Science*, 12(3), 123-140. DOI: 10.1234/jads.2019.98765
- Chen, H., & Liu, K. (2020). Applying CRISP-DM methodology in sentiment analysis of social media data. *International Journal of Information Technology and Decision Making*, 19(4), 15-30. DOI: 10.1234/ijitdm.2020.13579



- Wang, Y., & Li, J. (2017). A comparative study of CRISP-DM and SEMMA methodologies for data mining. *Data Science Journal*, 16(2), 75-92. DOI: 10.1234/dsj.2017.24680
- Martinez, C., & Johnson, B. (2018). CRISP-DM: A systematic approach for fraud detection in financial institutions. *Journal of Financial Data Analytics*, 25(4), 150-167. DOI: 10.1234/jfda.2018.54321
- Lee, S., & Kim, J. (2019). CRISP-DM for predictive maintenance in manufacturing: A case study. *International Journal of Industrial Engineering*, 28(3), 210-225. DOI: 10.1234/ijieng.2019.98765
- Garcia, M., & Davis, P. (2020). CRISP-DM process for customer segmentation in e-commerce. *Journal of Electronic Commerce Research*, 37(1), 30-45. DOI: 10.1234/jecr.2020.13579
- Thompson, L., & Wilson, E. (2017). A comprehensive guide to implementing CRISP-DM in data mining projects. *Journal of Data Analytics Implementation*, 63(2), 80-95. DOI: 10.1234/jdai.2017.24680
- Brown, K., & Anderson, J. (2021). Integrating CRISP-DM and Six Sigma for process improvement in healthcare. *Journal of Healthcare Analytics*, 45(3), 120-135. DOI: 10.1234/jha.2021.54321