



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA
MINERÍA DE DATOS



Proyecto Minería de Datos

Proyecto Selección de Carrera

Integrantes:

Luna Rodríguez Bryan Isaac

Ortega Mena Rafael Uriel

Pérez Saldaña Luis Mario

Rodríguez Dávalos Carolina

Grupo: 01

Semestre: 2024-1

Profesor: Gerardo Gabriel Carrasco Zúñiga

Proyecto Selección de Carrera.....	3
Introducción.....	3
Objetivo.....	3
Alcance del Proyecto.....	3
Planeación.....	4
1. Definir un plan de trabajo con estimación de fechas que considere las siguientes actividades:.....	4
1.1. Documentar el problema de Negocio.....	4
1.2. Establecer los elementos de Información requeridos.....	4
1.3. Aplicar métodos de limpieza y transformación de acuerdo a las características de la información.....	4
1.4. Presentar un reporte de estadísticas de la información ya transformada.....	4
1.5. Aplicar modelos o algoritmos de Minería de Datos.....	4
1.6. Identificando Clases, Reglas de Asociación, Reglas de Árboles de Decisión, Regresión y Agrupaciones.....	4
1.7. Documentar mediante reporte, la interpretación del resultado de cada algoritmo.....	4
Ejecución del proyecto.....	5
2. Ejecutar el plan, y documentar el resultado de cada actividad: Fecha planeada, fecha real. Estimado de tiempo invertido.....	5
2.1. Documentar el problema de Negocio.....	5
2.2. Establecer los elementos de Información requeridos.....	5
2.3. Aplicar métodos de limpieza y transformación de acuerdo a las características de la información.....	7
2.4. Presentar un reporte de estadísticas de la información ya transformada.....	9
2.5. Aplicar modelos o algoritmos de Minería de Datos. Identificando Clases, Reglas de Asociación, Reglas de Árboles de Decisión, Regresión y Agrupaciones.....	16
2.6. Documentar mediante reporte, la interpretación del resultado de cada algoritmo...	21
Ejecutando el plan de trabajo: fecha planeada vs fecha real.....	25
Propuesta de negocio.....	26
Business Case.....	27
Conclusión.....	28
Presentación.....	28

Proyecto Selección de Carrera

Introducción

La presente opción de proyecto se fundamenta en una base de información obtenida mediante la encuesta de elección de carrera, esta se diseñó con la finalidad de recopilar datos significativos de aspectos esenciales que repercuten en las decisiones académicas y profesionales de 310 participantes, los resultados obtenidos proporcionan una valiosa perspectiva sobre las preferencias y motivaciones de los encuestados.

La encuesta se estructura en dos secciones principales, en la primera parte, se solicitan datos demográficos clave, tales como nombre, edad actual, sexo, carrera y edad en la que se eligió la carrera profesional, estos datos proporcionan un contexto importante para comprender los perfiles individuales y los factores temporales asociados con la elección de una carrera.

En la segunda sección de la encuesta, esta se compone de preguntas estratégicas que evalúan la importancia relativa de diversos criterios y consideraciones en el proceso de elección de carrera, los participantes seleccionaron respuestas para estas preguntas mediante una escala de importancia, con el objetivo de proporcionar una medida cuantitativa de la relevancia de cada factor. Adicionalmente, en la segunda sección se incorporaron preguntas de opción múltiple y preguntas abiertas, a tal efecto de explorar de manera más detallada las preferencias percepciones y desafíos específicos que los encuestados experimentan al elegir una carrera profesional, la combinación de preguntas abiertas y de opción múltiple refuerza la comprensión de los matices y experiencias únicas de cada participante.

Con este enfoque global, nuestro proyecto busca exponer las dinámicas subyacentes que impulsan las elecciones de carrera, contribuyendo a la identificación de patrones, tendencias y áreas de interés que puedan ser de beneficio a instituciones educativas, tutores, orientadores académicos y en definitiva a los propios estudiantes, mediante la información recopilada de los datos recabados se proveerá de una visión profunda y valiosa que respalde la toma de decisiones informada en el ámbito educativo y profesional.

Objetivo

Aplicar los conocimientos adquiridos durante la materia de Minería de Datos

Alcance del Proyecto

- Definir un plan de trabajo con estimación de fechas
- Documentar el problema de Negocio
- Aplicar métodos de limpieza y transformación a los datos proporcionados
- Presentar un reporte estadístico
- Aplicar modelos o algoritmos de Minería de Datos
- Realizar documentación

Planeación

1. Definir un plan de trabajo con estimación de fechas que considere las siguientes actividades:

1.1. Documentar el problema de Negocio

Fecha estimada: 07 de Octubre de 2023

Fecha real: 07 de Octubre de 2023

1.2. Establecer los elementos de Información requeridos

Fecha estimada: 08 de Octubre de 2023

Fecha real: 08 de Octubre de 2023

1.3. Aplicar métodos de limpieza y transformación de acuerdo a las características de la información

Fecha estimada: 21 de Octubre de 2023

Fecha real: 04 de Noviembre de 2023

1.4. Presentar un reporte de estadísticas de la información ya transformada

Fecha estimada: 21 de Octubre de 2023

Fecha real: 11 de Noviembre de 2023

1.5. Aplicar modelos o algoritmos de Minería de Datos

Fecha estimada: 21 de Octubre de 2023

Fecha real: 11 de Noviembre de 2023

1.6. Identificando Clases, Reglas de Asociación, Reglas de Árboles de Decisión, Regresión y Agrupaciones

Fecha estimada: 22 de Octubre de 2023

Fecha real: 18 de Noviembre de 2023

1.7. Documentar mediante reporte, la interpretación del resultado de cada algoritmo

Fecha estimada: 22 de Octubre de 2023

Fecha real: 28 de Noviembre de 2023

Ejecución del proyecto

2. Ejecutar el plan, y documentar el resultado de cada actividad: Fecha planeada, fecha real. Estimado de tiempo invertido

2.1. Documentar el problema de Negocio

- 1) Reconocer los criterios de selección que más contribuyen a la decisión de carrera profesional de los estudiantes.
- 2) Conocer el grado de información y preparación de los estudiantes, y si cuentan con las capacidades necesarias en las materias que consideran relevantes previo a la toma de decisión de la carrera.
- 3) Identificar los puntos más importantes que estén relacionados con la satisfacción y pasión por la carrera. Descubrir si es una cuestión de nivel de preparación, nivel de sus habilidades, sexo, si se trató de una decisión personal, y/o edad de la toma de decisión.
- 4) Identificar, con la información del punto 2.1, inciso 3, las áreas de oportunidad de un estudiante de ingeniería para poder tener una mejor experiencia/desempeño durante la carrera.
- 5) Predecir, con la información del punto 2.1, inciso 3, si un estudiante estaría satisfecho escogiendo una ingeniería como su carrera profesional.

2.2. Establecer los elementos de Información requeridos

1. Las respuestas a las siguientes preguntas contenidas en la encuesta son relevantes para contestar este punto:

¿Consideraste los recursos financieros y materiales que requiere tu carrera profesional para elegir?
¿Consideraste el Pase automático reglamentado como un criterio de decisión de carrera profesional?
¿Consideraste el promedio de calificación mínimo para elegir tu carrera profesional?
¿Consideraste la distancia a la universidad como criterio de selección de tu carrera profesional?
¿El prestigio del título profesional fue factor para elegir tu carrera profesional?
¿El salario promedio para profesionistas de tu carrera fue criterio de decisión?

2. Las respuestas a las siguientes preguntas contenidas en la encuesta son relevantes para contestar este punto:

¿Investigaste las oportunidades laborales de tu carrera profesional?
¿Visitaste alguna feria o expo de Carreras Profesionales para documentarte?
¿Qué materias del bachillerato consideramos relevantes en la elección de tu carrera profesional?
¿Qué calificación obtuviste en las materias mencionadas? (Del 1 al 10)
Menciona 5 habilidades que requiere tu carrera
¿Cómo te calificas en cada habilidad? (Del 1 al 10, siendo 10 la más alta)

3. Las respuestas a las siguientes preguntas contenidas en la encuesta son relevantes para ser tomadas en consideración como variables independientes del modelo:

a. Nivel de Preparación (valor numérico del 0 al 5)

¿Investigaste las oportunidades laborales de tu carrera profesional?
¿Visitaste alguna feria o expo de Carreras Profesionales para documentarte?
¿Qué materias del bachillerato consideramos relevantes en la elección de tu carrera profesional?
¿Qué calificación obtuviste en las materias mencionadas? (Del 1 al 10)

b. Nivel de Habilidades (valor numérico del 0 al 5)

Menciona 5 habilidades que requiere tu carrera
¿Cómo te calificas en cada habilidad? (Del 1 al 10, siendo 10 la más alta)

c. Tipo de Decisión (Personal, Familiares, Amigos, Otras)

La elección de tu carrera profesional, ¿fue una decisión?

d. Sexo (M o F)

Sexo:

e. Edad de la Toma de Decisión (rangos)

Edad en que Elegiste tu Carrera Profesional:
--

Las respuesta a las siguiente pregunta es relevantes para ser tomadas en consideración como variables dependientes del modelo:

a. Nivel de Pasión y Cumplimiento de Expectativas (valor numérico del 0 al 5)

¿Tu carrera profesional es lo que realmente querías?
--

¿Te apasiona tu carrera profesional?

Se hará un modelo de correlación para identificar la relación entre cada una de las variables independientes y la variable dependiente. Mientras más fuerte sea la correlación, se podrá deducir cuáles son los aspectos más importantes que estén relacionados con la satisfacción y pasión por la carrera.

4. Identificados los aspectos más importantes en el punto 2.2, inciso 3, si existe correlación alguna entre los puntos propuestos, se tratará de explicar el por qué de los distintos niveles de pasión por las carreras de ingeniería ,y se propondrán algunas áreas de oportunidad para mejorar el desempeño y cumplimiento de expectativas de los estudiantes.

2.3. Aplicar métodos de limpieza y transformación de acuerdo a las características de la información

Para la columna de CARRERA se aplicó transformación a los datos, con el objetivo de que todas las palabras estuvieran en minúscula, se completaron palabras, se corrigieron las faltas de ortografía y se quitaron espacios extras.

Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería Eléctrica Electrónica	ingeniería eléctrica electrónica
Ingeniería Industrial	ingeniería industrial
Ingeniería en Computación	ingeniería en computación
Ingeniería Mecatrónica	ingeniería mecatrónica
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería Mecánica	ingeniería mecánica
Ingeniería Geofísica	ingeniería geofísica
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación
Ingeniería en Computación	ingeniería en computación

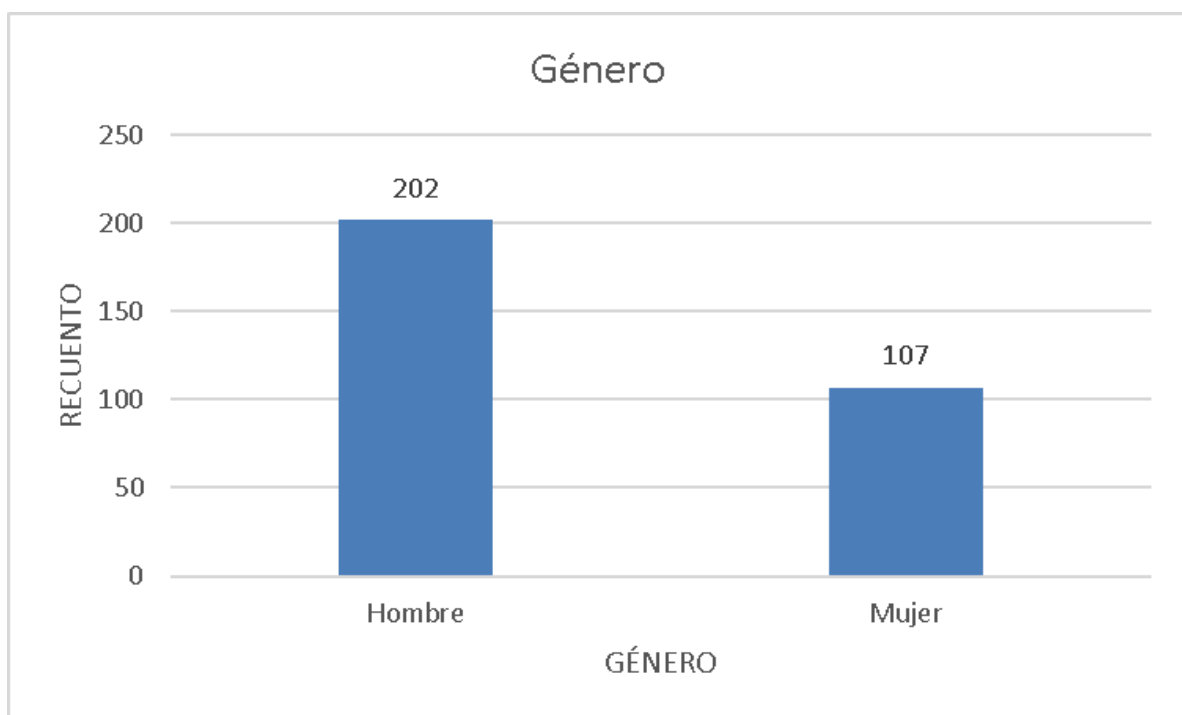
Para la columna GENERO, se aplicó transformación a los datos, con el objetivo de que todas las palabras estuvieran en minúscula.

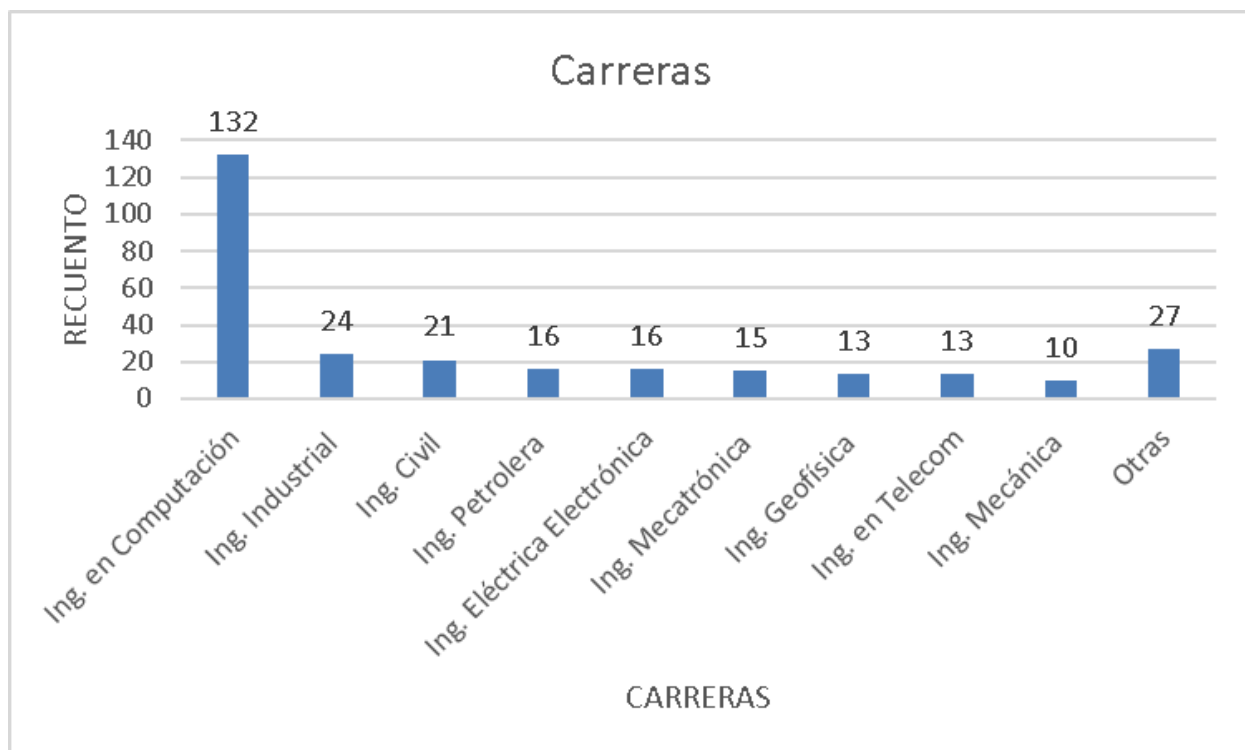
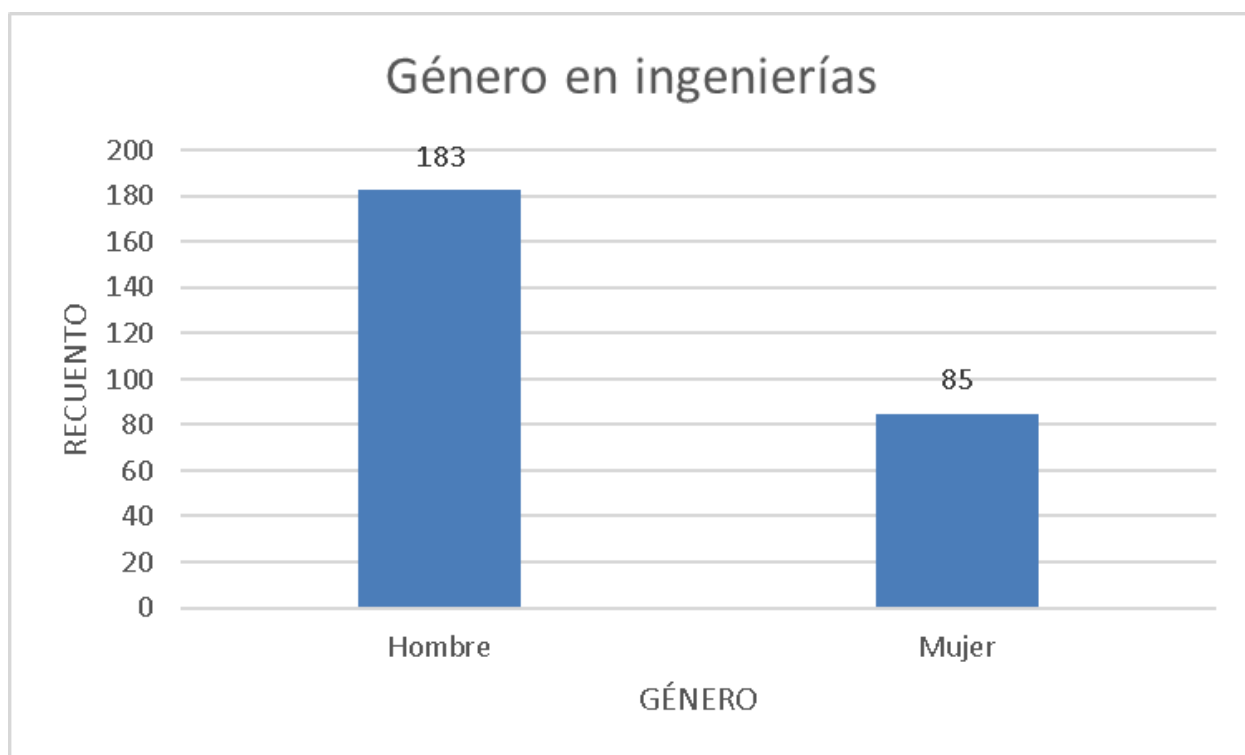
Hombre	hombre
Hombre	hombre
Hombre	hombre
Mujer	mujer
Mujer	mujer
Mujer	mujer
Hombre	hombre
Hombre	hombre
Hombre	hombre
Mujer	mujer
Hombre	hombre
Hombre	hombre

Para la comuna EDAD_ELECCION_CARRERA se comprobó que fuera menor o igual a la columna EDAD.

22	18	1
21	18	1
20	15	1
21	18	1
21	17	1
20	18	1
19	17	1

2.4. Presentar un reporte de estadísticas de la información ya transformada



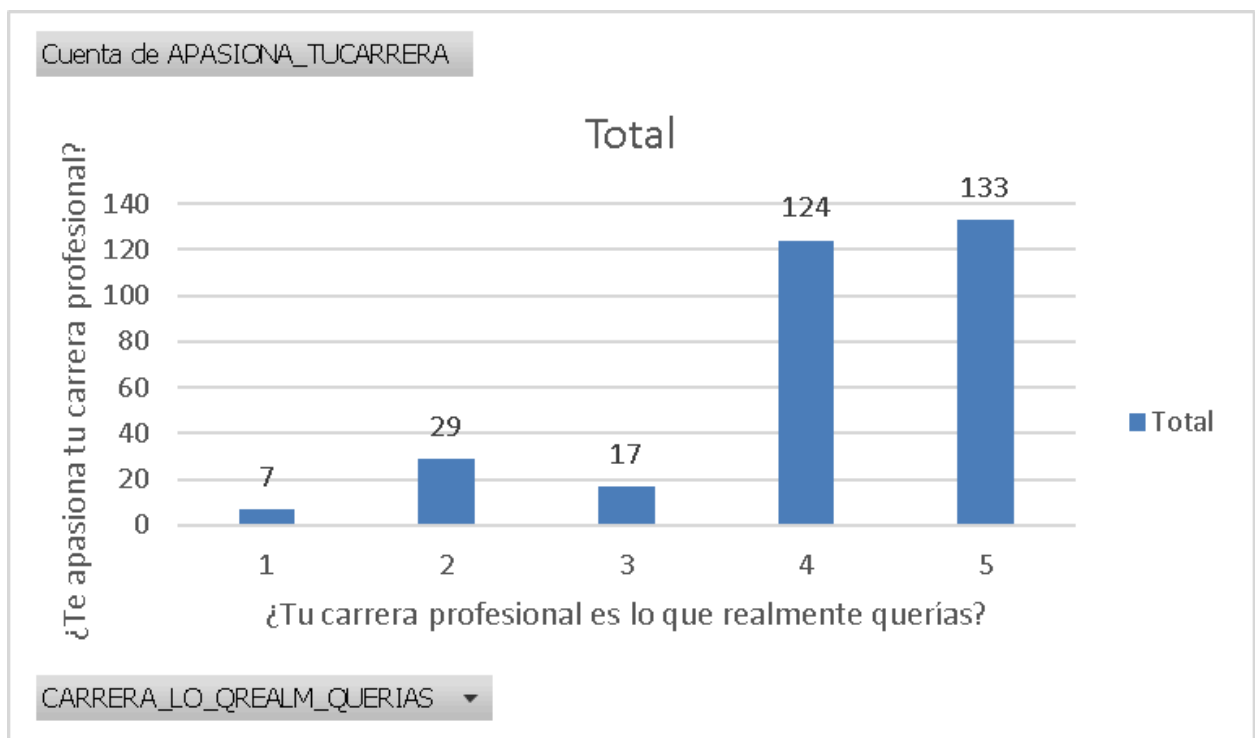


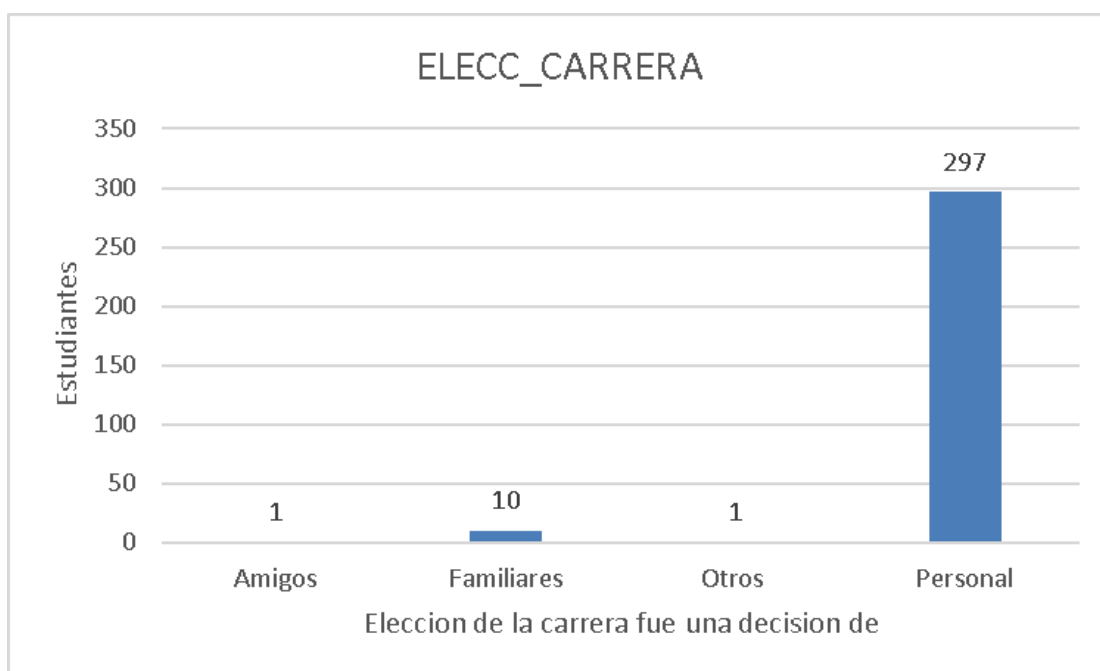
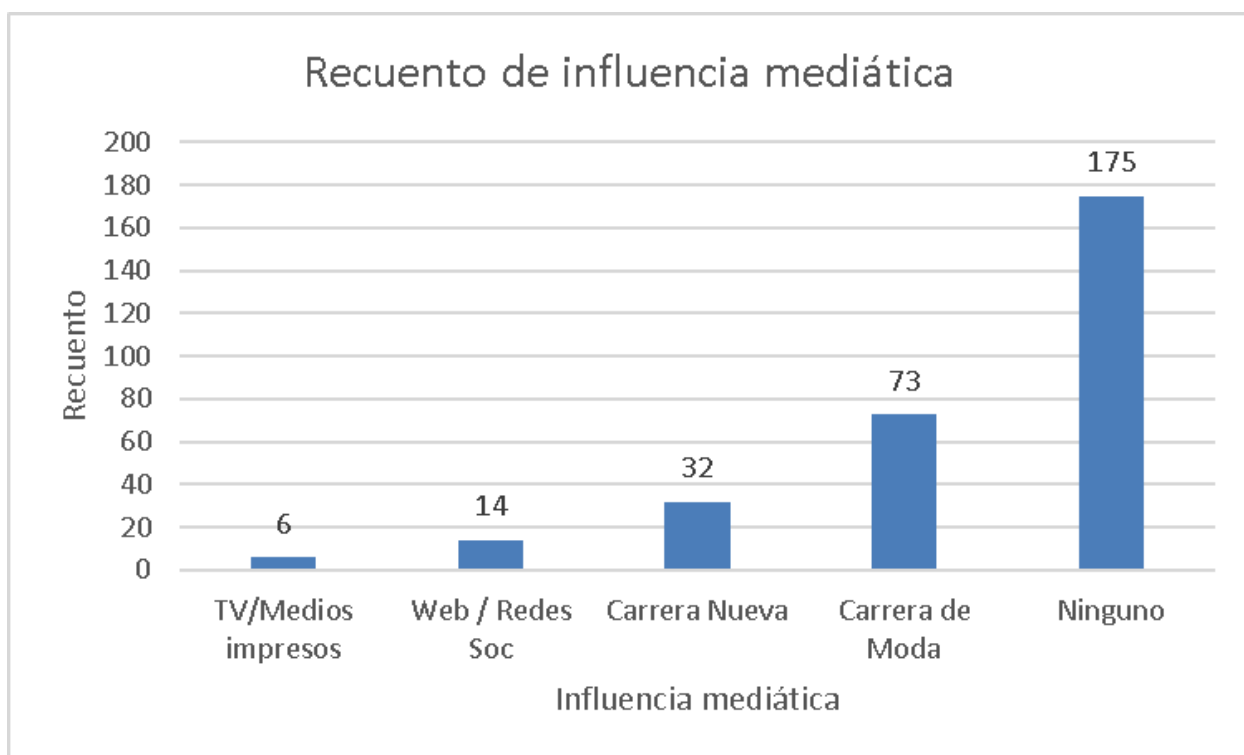
Otras se refiere a: Administración (5), Diseño y comunicación visual(4), Contaduría (4), Medicina (4), Arquitectura (3), Informática (3), Ingeniería geológica (3), Actuaría (2), Ingeniería en minas y metalurgia(2), Trabajo social (2), Derecho (2), Gastronomía (1), Ingeniería biomédica (1), Antropología social (1), Ingeniería en sistemas biomédicos(1),

Ciencia política (1), Ingeniería geomática (1), Economía (1), Ingeniería química (1), Sociología (1), Matemáticas (1), Ciencias de la comunicación(1), Matemáticas aplicadas (1), Biología (1), Medicina veterinaria (1), Fisioterapia (1), Química farmacéutica biológica(1).

Valores	
Suma de APASIONA_TUCARRERA	1308
Suma de CARRERA_LO_QREALM_QUERIAS	1277
Suma de SALARIO_PROM	981
Suma de PROM_CALIF_MIN	977
Suma de REC_FINANCIEROS	952
Suma de PASE_AUTOMATICO	938
Suma de PRESTIG_TITULO	935
Suma de DIST_ALA_UNIV	733

Por lo que podemos ver que lo que la gente considera que tiene mayor importancia es que le apasiona su carrera y que realmente es lo que quieren estudiar, lo siguiente es el salario promedio de su carrera y el promedio mínimo, y a lo que le dan menos importancia es al prestigio del título y con una diferencia considerable la distancia a la universidad.



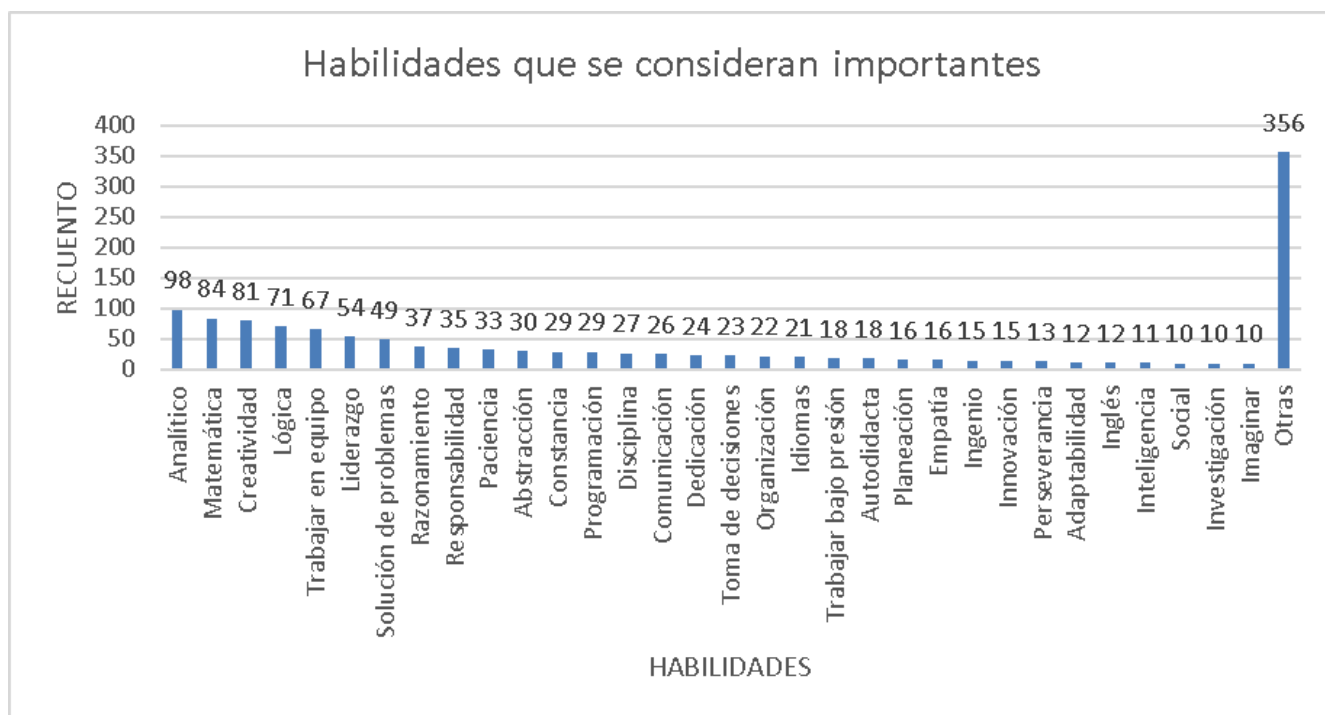


2. Primero verificamos cuales son las qué materias del bachillerato consideran relevantes en la elección de tu carrera profesional



Observamos que las materias que consideran más importantes los encuestados son: Física, Matemáticas, Cálculo, Química, Computación . El apartado de Otras, son materias que en el recuento no tienen un valor mayor a 4.

Otras se refiere a: Bases de datos(4), Español(4), Lengua Española(4), Orientación educativa(4), Idiomas(4), Dibujo Constructivo(3), Ciencias Sociales(3), Redes(3), Latín(3), Etimologías Grecolatinas(2), Inteligencia Artificial(2), Historia Universal(2), Electrónica(2), Diseño(2), Lectura(2), Ciencias políticas y sociales(2), Etimológicas(2), Estética(2), Estadística(2), Historia de México(2), Teatro(2), Robótica(2), Ciencias de la salud(2), Dibujo técnico(2), Mecánica(2), Morfofisiología(2), Geología(2), Problemas Políticos(1), Termodinámica(1), Geografía económica(1), Diseño Web(1), Ortografía(1), Idioma(1), Cad(1), Dispositivos digitales(1), Sistemas operativos(1), Anatomía(1), Diseño Digital(1), E-Learning(1), Poo(1), Electricidad y Magnetismo(1), Procesamiento de Lenguaje Natural(1), Italiano(1), Recursos humanos(1), Antropología(1), Geología Y Mineralogía(1), Electrónica digital(1), Ciencias(1), Artes(1), Toma de decisiones(1), Lengua extranjera(1), Orientación juvenil(1), Lenguaje(1), Pintura(1), Lenguaje y Comunicación(1), Administración de proyectos de software(1), Lenguajes Formales y Automátas(1), Problemas sociales políticos y económicos de México(1), Cinemática y dinámica(1), Bioquímica(1), Circuitos digitales(1), Geoeconomía(1), Manufactura(1), Algoritmos(1), Marketing(1), Análisis de Sistemas y Señales(1), Circuitos Lógicos Combinatorios(1), Salud pública(1), Circuitos Lógicos Secuenciales(1), Sociales y económicos de México(1), Metodología(1), Diseño ambiental(1), Artes plásticas(1), TIC(1), Morfología humana(1), Tutoría(1), Negocios Electrónicos(1), Admisnitración de proyectos de software(1).



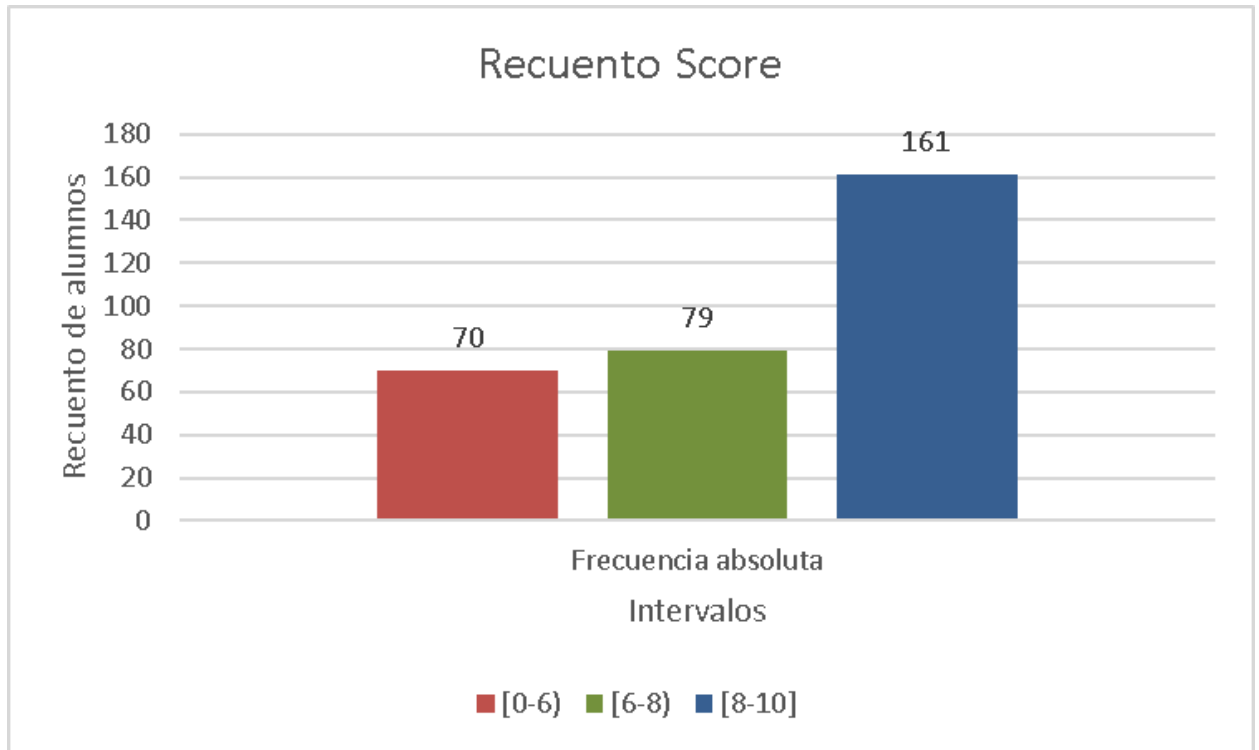
Observamos que las habilidades que consideran más importantes los encuestados son: Analítico, Matemática, Creatividad, Lógica, Liderazgo . El apartado de Otras, son materias que en el recuento no tienen un valor mayor a 9.

Otras se refiere a: Física(9), Observación(9), Destreza(9), Diseño(9), Tolerancia(7), Comprensión(7), Visión(7), Proactivo(6), Memoria(6), Síntesis(6), Eficiencia(5), Negociar(5), Computación(5), Compromiso(5), Honestidad(5), Expresión(4), Espacial(4), Cálculo Mental(4), Control(4), Ética(4), Emprendedor(4), Redacción(4), Interpretación(4), Lectura(4), Iniciativa(4), Práctica(3), Conocimiento(3), Radical(3), Administración(3), Confianza(3), Extrovertido(3), Comprensión de textos(3), Entendimiento(3), Rapidez(3), Objetivo(3), Estrategia(3), Pasión(3), Atención(3), Actitud(3), Agilidad mental(2), Buena ortografía(2), Desempeño(2), Dibujo(2), Intuición(2), Manejo de equipo(2), Aplicación(2), Estudiar(2), Respeto(2), Mental(2), Tenacidad(2), Dinámico(2), Precisión(2), Curiosidad(2), Imaginación espacial(2), Compañerismo(2), Razonamiento lógico(2), Optimización(2), Resiliencia(2), Concentración(2), Sentido Común(2), Pensamiento Crítico(2), Determinación(2), Pensar(2), Teoría(2), Flexibilidad(2), Electrónica(2), Asertividad(2), Primeros auxilios(1), Factibilidad de ideas(1), Exposición de temas(1), Generar(1), Estadística(1), Geohistoria(1), Análisis de datos cuantitativos y cualitativos(1), Geología(1), Respuestas rápidas(1), Geometría(1), Tiempo(1), Hablar en público(1), Estructurar(1), Hidráulica(1), Procesos(1), Diligencia(1), Rapidez mental(1), Contabilidad(1), Detección(1), Administración del tiempo(1), Sensibilidad(1), Capacidad(1), Soporte físico(1), Implementar(1), Pensamiento flexible(1), Improvisación(1), Aptitud numérica(1),

Influyente(1), Conocimientos de normas(1), Informática(1), Probabilística(1), Aprendizaje(1), Productivo(1), Capacidad de desarrollo(1), Rápida adaptación(1), Diseñar estructuras(1), Experimentación(1), Condición física(1), Reflexionar(1), Verbal(1), Actualización(1), Visión del entorno(1), Saber buscar información(1), Visualizar espacio(1), Facilidad para describir objetos o lugares(1), Análisis Espacial(1), Software(1), Análisis financiero(1), Campo(1), Trabajo individual(1), G. P. S.(1), Ubicación(1), Pensamiento prospectivo(1), Disposición(1), Conocimiento del contexto ambiental(1), Docencia(1), Plantear problemas(1), Documentar(1), Pragmática(1), Cognitiva(1), Prevención(1), Dominio conocimientos(1), Destrezas manuales para procedimientos específicos(1), Álgebra(1), Procesamiento de datos(1), Algorítmica(1), Producción(1), Creación de proyectos(1), Profesionalismo(1), Manejo de información(1), Ética profesional(1), Manejo de señales(1), Excelencia(1), Manejo de sistemas operativos(1), Análisis de riesgos(1), Altruismo(1), Recto(1), Crear(1), Redes de Datos(1), Apasionado(1), Relaciones interpersonales(1), Meticulosa(1), Resolver casos en un tiempo establecido(1), Metodología(1), Análisis de sistemas(1), Modelado(1), Retención(1), Motriz(1), Seguridad(1), Activo(1), Sensibilidad de espacios(1), Enseñar(1), Ser estructurado(1), Conceptualización(1), Financiera(1), Obtener información(1), Análisis e Investigación(1), Entrega(1), Tecnología(1), Amabilidad(1), Construcción(1), Análisis Crítico(1), Fuerza física(1), Esfuerzo(1), Desarrollo(1), Certificaciones(1), Trabajo(1), Inteligencia emocional(1), Trato con personas(1), Interdisciplina(1), Ventas(1), Interés(1), Interés por la tierra(1), Instalación(1), Visualizar a largo plazo(1), Integrar conceptos(1), Astucia(1), Integridad(1).

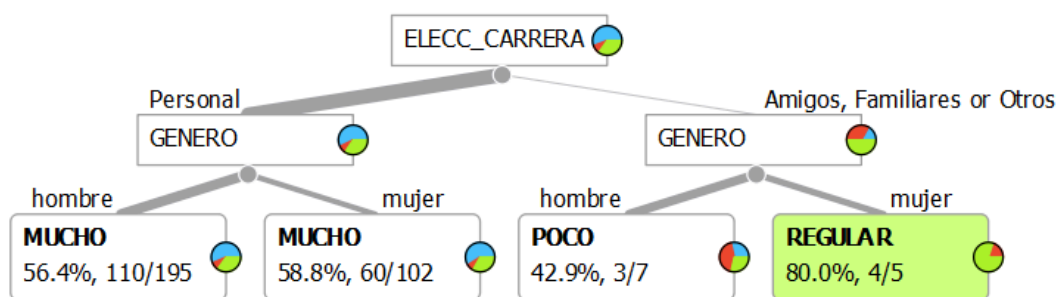
Se realizó el promedio de las calificaciones y habilidades, que nos permitirán asignarles un score, con el cual problemas identificar tres clases,

- [0-6]=Malo
- [6-8]=Regular
- [8-10]=Bueno



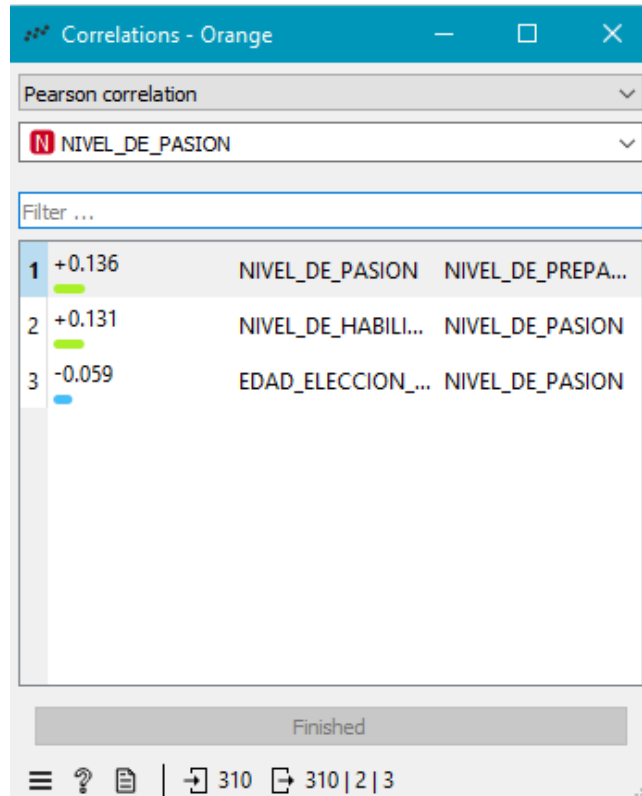
2.5. Aplicar modelos o algoritmos de Minería de Datos. Identificando Clases, Reglas de Asociación, Reglas de Árboles de Decisión, Regresión y Agrupaciones

1. Árboles de Decisión



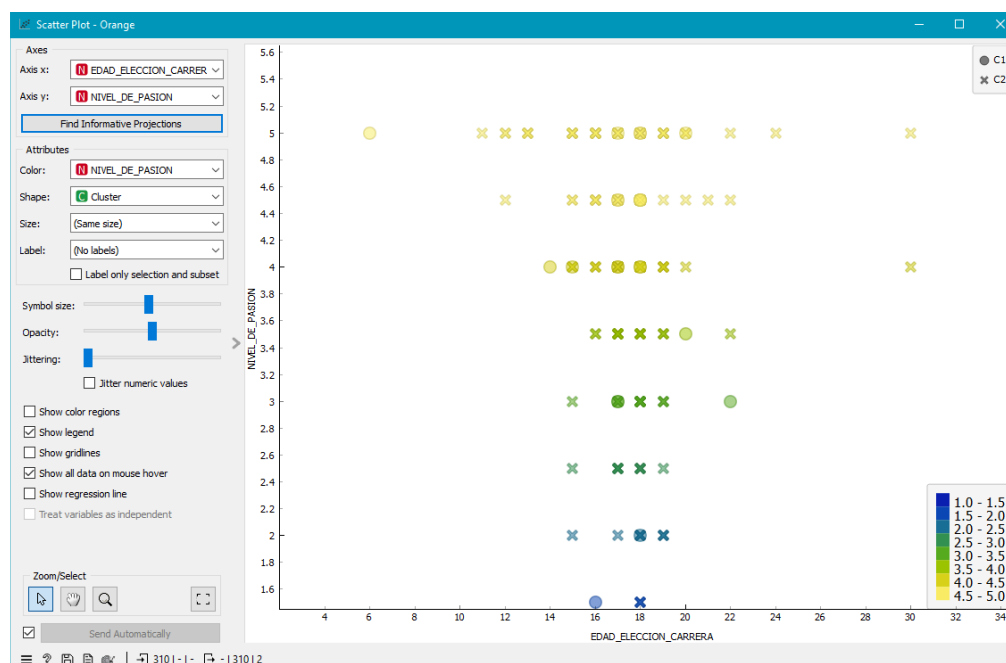
- Variable Dependiente: Grado de Pasión (Mucho, Regular, Poco)
- Variables Independientes: Elección de Carrera (Personal, Otros), Género (Hombre, Mujer)

2. Correlación

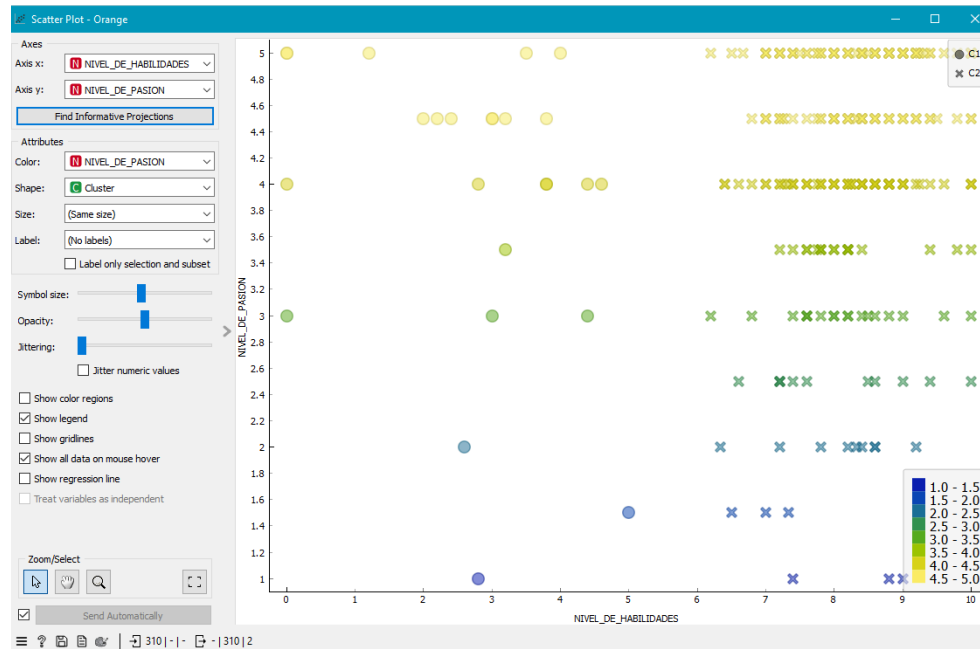


- Variable Dependiente: Nivel de Pasión, Rango [0-5]

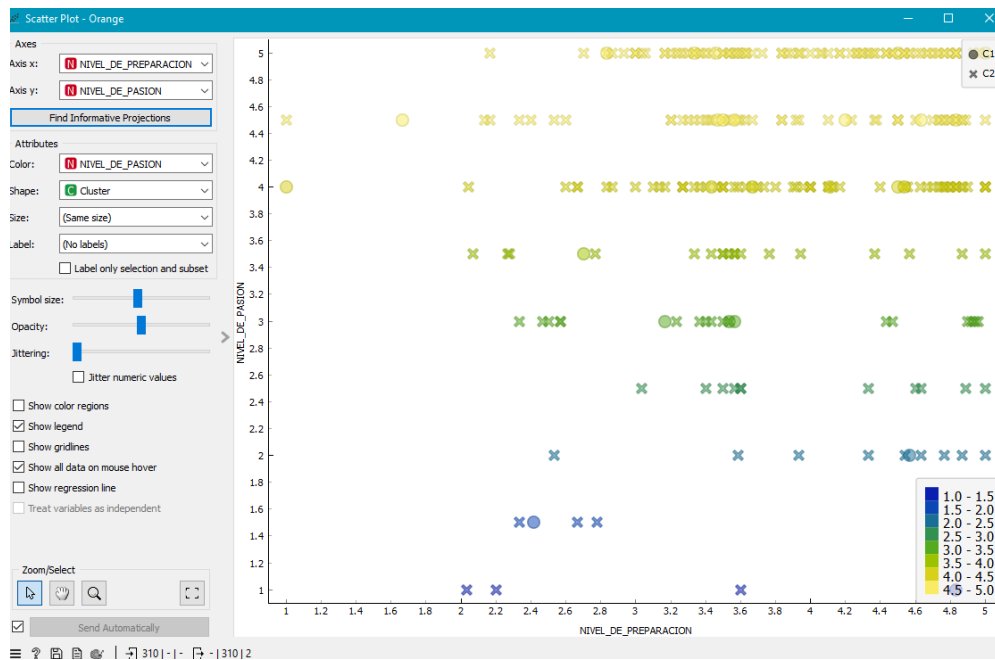
3. Agrupaciones



- Variable Dependiente: Nivel de Pasión, Rango [0-5]
- Variables Independiente: Edad Elección de Carrera, Rango [0-30]



- Variable Dependiente: Nivel de Pasión, Rango [0-5]
- Variables Independiente: Nivel de Habilidades, Rango [0-5]



- Variable Dependiente: Nivel de Pasión, Rango [0-5]
- Variables Independiente: Nivel de Preparación, Rango [0-5]

4. Reglas de Asociación

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	
0.113	0.875	0.129	5.075	1.336	0.028	INFLU_MEDIATICA=Web / Redes Soc	→ GENERO=hombre
0.081	0.962	0.084	7.808	1.468	0.026	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc	→ GENERO=hombre
0.071	0.917	0.077	8.458	1.400	0.020	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc	→ GENERO=hombre
0.068	0.875	0.077	6.458	1.750	0.029	CARRERA=Ingeniería en Computación, CARR_PROF_PADRES=Sin Carrera	→ TIPOS_ESPACIOS=Oficinas
0.068	0.875	0.077	8.458	1.336	0.017	INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.058	0.947	0.061	10.684	1.447	0.018	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc	→ GENERO=hombre
0.052	0.889	0.058	8.611	1.778	0.023	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales	→ TIPOS_ESPACIOS=Oficinas
0.048	0.882	0.055	9.118	1.765	0.021	CARRERA=Ingeniería en Computación, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Sin Carrera	→ TIPOS_ESPACIOS=Oficinas
0.048	0.938	0.052	12.688	1.432	0.015	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.045	1.000	0.045	11.071	2.000	0.023	GENERO=mujer, CARRERA=Ingeniería en Computación, GRADO_DE_PASION=REGULAR	→ TIPOS_ESPACIOS=Oficinas
0.045	0.875	0.052	9.688	1.750	0.019	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Ninguna, GRADO_DE_PASION=MUCHO	→ TIPOS_ESPACIOS=Oficinas
0.045	0.875	0.052	12.688	1.336	0.011	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, CARR_PROF_PADRES=S/carrera	→ GENERO=hombre
0.045	0.933	0.048	13.533	1.425	0.013	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→ GENERO=hombre
0.045	0.875	0.052	12.688	1.336	0.011	INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.042	1.000	0.042	15.615	1.527	0.014	INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=REGULAR	→ GENERO=hombre
0.042	0.867	0.048	13.533	1.323	0.010	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.042	0.867	0.048	8.800	2.035	0.021	GENERO=hombre, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Sociales	→ CARRERA=Ingeniería en Computación
0.042	0.867	0.048	10.333	1.733	0.018	GENERO=hombre, CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales	→ TIPOS_ESPACIOS=Oficinas
0.039	0.857	0.045	11.071	1.714	0.016	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Ninguna, CARR_PROF_PADRES=Sin Carrera	→ TIPOS_ESPACIOS=Oficinas
0.039	0.857	0.045	11.071	1.714	0.016	GENERO=hombre, CARRERA=Ingeniería en Computación, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Sin Carrera	→ TIPOS_ESPACIOS=Oficinas
0.039	0.923	0.042	15.615	1.410	0.011	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→ GENERO=hombre
0.039	0.923	0.042	15.615	1.410	0.011	INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=S/carrera	→ GENERO=hombre
0.035	0.917	0.039	16.917	1.400	0.010	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.035	0.917	0.039	16.917	1.400	0.010	TIPOS_ESPACIOS=Otros	→ GENERO=hombre
0.032	0.909	0.035	9.818	2.609	0.020	GENERO=mujer, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=S/carrera	→ INFLU_MEDIATICA=Ninguno
0.032	0.909	0.035	15.727	1.629	0.012	GENERO=mujer, INFLU_MEDIATICA=Ninguna, CARR_PROF_PADRES=Sin Carrera	→ GRADO_DE_PASION=MUCHO
0.032	0.909	0.035	18.455	1.388	0.009	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→ GENERO=hombre
0.032	0.909	0.035	18.455	1.388	0.009	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Diferente	→ GENERO=hombre
0.032	1.000	0.032	20.300	1.527	0.011	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=S/carrera	→ GENERO=hombre
0.032	0.909	0.035	14.091	1.818	0.015	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales, CARR_PROF_PADRES=Diferente	→ TIPOS_ESPACIOS=Oficinas

5. Identificación de Clases

Predictions - Orange

Show probabilities for

Classes in data

Show classification errors

	CN2 rule inducer	error	Naive Bayes	error
1	0.12 : 0.75 : 0.12 → MALO	0.250	0.00 : 1.00 : 0.00 → MALO	0.000
2	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.00 : 1.00 → REGU...	0.002
3	0.17 : 0.67 : 0.17 → MALO	0.333	0.00 : 0.93 : 0.07 → MALO	0.074
4	0.20 : 0.60 : 0.20 → MALO	0.400	0.00 : 0.99 : 0.01 → MALO	0.012
5	0.12 : 0.75 : 0.12 → MALO	0.250	0.00 : 1.00 : 0.00 → MALO	0.001
6	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.00 : 1.00 → REGU...	0.000
7	0.12 : 0.75 : 0.12 → MALO	0.250	0.00 : 0.93 : 0.07 → MALO	0.070
8	0.99 : 0.01 : 0.01 → BUENO	0.012	1.00 : 0.00 : 0.00 → BUENO	0.000
9	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.00 : 1.00 → REGU...	0.000
10	0.12 : 0.75 : 0.12 → MALO	0.250	0.00 : 0.99 : 0.01 → MALO	0.007
11	0.99 : 0.01 : 0.01 → BUENO	0.012	1.00 : 0.00 : 0.00 → BUENO	0.000
12	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.87 : 0.13 → MALO	0.868
13	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.03 : 0.97 → REGU...	0.031
14	0.12 : 0.75 : 0.12 → MALO	0.250	0.00 : 1.00 : 0.00 → MALO	0.003
15	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.00 : 1.00 → REGU...	0.001
16	0.01 : 0.01 : 0.98 → REGU...	0.024	0.03 : 0.00 : 0.97 → REGU...	0.026
17	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.01 : 0.99 → REGU...	0.008
18	0.99 : 0.01 : 0.01 → BUENO	0.012	1.00 : 0.00 : 0.00 → BUENO	0.002
19	0.99 : 0.01 : 0.01 → BUENO	0.012	1.00 : 0.00 : 0.00 → BUENO	0.005
20	0.99 : 0.01 : 0.01 → BUENO	0.012	1.00 : 0.00 : 0.00 → BUENO	0.000
21	0.01 : 0.01 : 0.98 → REGU...	0.024	0.00 : 0.00 : 1.00 → REGU...	0.004

Show performance scores

Target class: (Average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
CN2 rule inducer	1.000	1.000	1.000	1.000	1.000	1.000
Naive Bayes	0.994	0.945	0.946	0.948	0.945	0.913

311

1

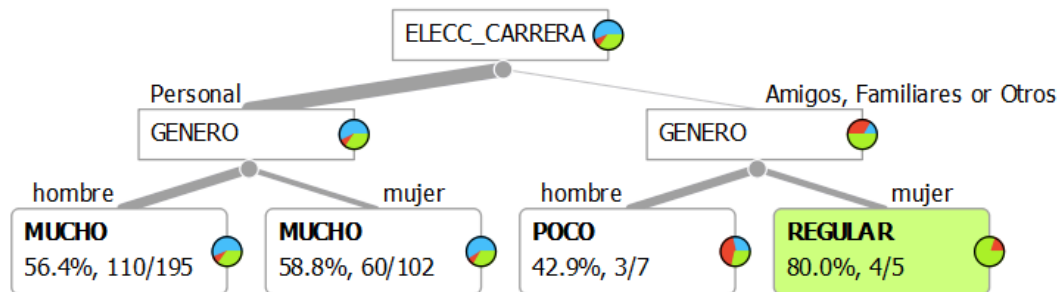
2x310

	CN2 rule inducer	error	Naive Bayes	error	tipo_ES1
291	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.98 : 0.00 : 0.02 → BUENO</u>	<u>0.017</u>	BUENO
292	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.98 : 0.00 : 0.02 → BUENO</u>	<u>0.020</u>	BUENO
293	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.98 : 0.00 : 0.02 → BUENO</u>	<u>0.021</u>	BUENO
294	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.98 : 0.00 : 0.02 → BUENO</u>	<u>0.024</u>	BUENO
295	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.97 : 0.00 : 0.03 → BUENO</u>	<u>0.028</u>	BUENO
296	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.97 : 0.00 : 0.03 → BUENO</u>	<u>0.030</u>	BUENO
297	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.97 : 0.00 : 0.03 → BUENO</u>	<u>0.031</u>	BUENO
298	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.96 : 0.00 : 0.04 → BUENO</u>	<u>0.044</u>	BUENO
299	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.95 : 0.00 : 0.05 → BUENO</u>	<u>0.053</u>	BUENO
300	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.85 : 0.00 : 0.15 → BUENO</u>	<u>0.149</u>	BUENO
301	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.82 : 0.00 : 0.18 → BUENO</u>	<u>0.183</u>	BUENO
302	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.79 : 0.00 : 0.21 → BUENO</u>	<u>0.211</u>	BUENO
303	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.78 : 0.00 : 0.22 → BUENO</u>	<u>0.221</u>	BUENO
304	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.73 : 0.00 : 0.27 → BUENO</u>	<u>0.266</u>	BUENO
305	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.09 : 0.00 : 0.91 → REGU...</u>	<u>0.910</u>	BUENO
306	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.03 : 0.00 : 0.97 → REGU...</u>	<u>0.966</u>	BUENO
307	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.996</u>	BUENO
308	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.998</u>	BUENO
309	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.999</u>	BUENO
310	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.999</u>	BUENO
311	<u>0.99 : 0.01 : 0.01 → BUENO</u>	<u>0.012</u>	<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.999</u>	BUENO

Naive Bayes	error		CN2 rule inducer	error
<u>0.00 : 0.04 : 0.50 → MALO</u>	<u>0.550</u>	291	<u>0.17 : 0.07 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.63 : 0.37 → MALO</u>	<u>0.374</u>	292	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.47 : 0.53 → REGU...</u>	<u>0.469</u>	293	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.47 : 0.00 : 0.53 → REGU...</u>	<u>0.533</u>	294	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.62 : 0.38 → MALO</u>	<u>0.618</u>	295	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.37 : 0.63 → REGU...</u>	<u>0.633</u>	296	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.21 : 0.00 : 0.79 → REGU...</u>	<u>0.788</u>	297	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.81 : 0.19 → MALO</u>	<u>0.807</u>	298	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.83 : 0.17 → MALO</u>	<u>0.833</u>	299	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.87 : 0.13 → MALO</u>	<u>0.868</u>	300	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.96 : 0.04 → MALO</u>	<u>0.958</u>	301	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.96 : 0.04 → MALO</u>	<u>0.962</u>	302	<u>0.17 : 0.67 : 0.17 → MALO</u>	<u>0.333</u>
<u>0.00 : 0.97 : 0.03 → MALO</u>	<u>0.974</u>	303	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.02 : 0.00 : 0.97 → REGU...</u>	<u>0.977</u>	304	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.02 : 0.00 : 0.98 → REGU...</u>	<u>0.978</u>	305	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.02 : 0.00 : 0.98 → REGU...</u>	<u>0.983</u>	306	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.01 : 0.00 : 0.99 → REGU...</u>	<u>0.994</u>	307	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.01 : 0.00 : 0.99 → REGU...</u>	<u>0.994</u>	308	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.995</u>	309	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>0.00 : 0.00 : 1.00 → REGU...</u>	<u>0.997</u>	310	<u>0.20 : 0.60 : 0.20 → MALO</u>	<u>0.400</u>
<u>1.00 : 0.00 : 0.00 → BUENO</u>		311	<u>0.01 : 0.01 : 0.98 → REGU...</u>	

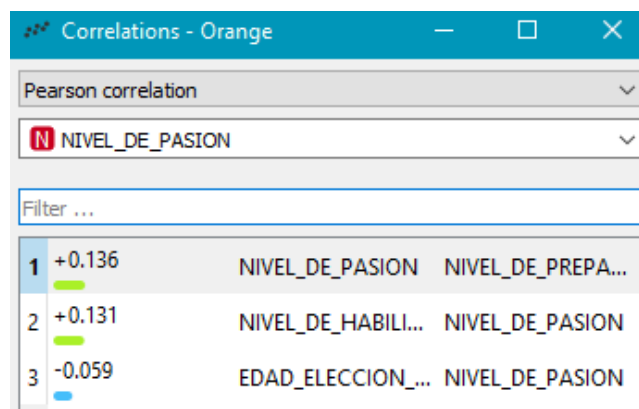
2.6. Documentar mediante reporte, la interpretación del resultado de cada algoritmo

1. Árboles de Decisión



En el árbol de decisión se observa una marcada relación entre quién toma la decisión de carrera y el grado de pasión y gusto por la misma. Si la decisión es personal, es más probable que sea MUCHO el grado de pasión por la carrera, mientras que si es una decisión tomada por familiares o amigos, la tendencia indica que el grado de pasión por la carrera sea REGULAR o POCO. La tendencia también se acentúa ligeramente más en hombres, ya que muestran un grado de interés todavía más bajo que las mujeres cuando no son ellos quienes deciden. Sin embargo, entre hombres y mujeres que deciden estudiar la carrera por cuenta propia, no se observa alguna disparidad notable en el grado de pasión por la carrera, y por lo tanto, el grado de pasión por las carreras encuestadas, que son en su mayoría ingenierías, no es una cuestión de género. De este algoritmo, podemos deducir que es importante que uno mismo sea quien decida lo que estudiar.

2. Correlación

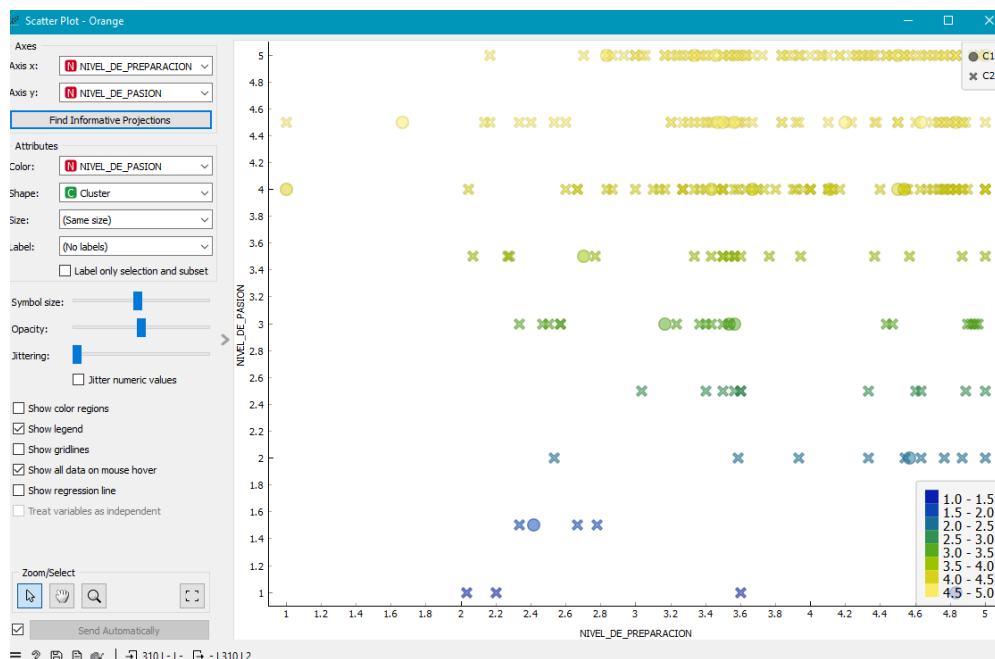


Se buscó saber si el nivel de preparación de los estudiantes previo a ingresar a la universidad, el nivel en las habilidades que consideran relevantes para sus estudios, o la edad de elección de su carrera tienen relación alguna con el nivel de pasión por la misma. Se tomó el nivel de pasión como variable dependiente de cada uno de los tres factores recién mencionados. El algoritmo de Pearson indica que la correlación entre los tres

criterios y el nivel de pasión es muy débil, y no arroja nada concluyente para poder asegurar que alguno de estos puntos es crucial para que un estudiante disfrute o no de su carrera. Los valores se encuentran alrededor del 0, lo que significa que entre las variables no hay ni correlación positiva, ni correlación negativa considerables.

Si se observan los datos, tampoco se puede distinguir ninguna relación. Los datos incluyen alumnos con un nivel de preparación pobre (malas calificaciones en preparatoria, y poca investigación sobre la carrera), que indican que la carrera ha cumplido con sus expectativas y muestran pasión por la misma. Lo mismo ocurre con estudiantes que tomaron su decisión de carrera a una edad avanzada, y aún así disfrutaban de ella.

3. Agrupación



Si existiera una correlación entre el nivel de preparación y el nivel de pasión por la carrera, se esperaría que el resultado del algoritmo de clustering por k-medias agrupara a los estudiantes de forma que los más preparados tuvieran un grado más alto de pasión, y lo contrario con los menos preparados.

Pero como ya se observó en los resultados de correlación, no existe tal indicativo y por lo tanto, la agrupación se da de una forma muy extraña. Lo mismo ocurre con los otros dos factores, como se observa en las imágenes de los resultados de este algoritmo.

En el Scatter Plot, se observa claramente como alumnos con un grado de preparación relativamente bajo, pueden disfrutar de la carrera, así como alumnos quienes con un grado alto de preparación, no encuentran pasión por sus estudios.

4. Reglas de Asociación

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		
0.113	0.875	0.129	5.075	1.336	0.028	INFLU_MEDIATICA=Web / Redes Soc	→	GENERO=hombre
0.081	0.962	0.084	7.808	1.468	0.026	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc	→	GENERO=hombre
0.071	0.917	0.077	8.458	1.400	0.020	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc	→	GENERO=hombre
0.068	0.875	0.077	6.458	1.750	0.029	CARRERA=Ingeniería en Computación, CARR_PROF_PADRES=Sin Carrera	→	TIPOS_ESPACIOS=Oficinas
0.068	0.875	0.077	8.458	1.336	0.017	INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.058	0.947	0.061	10.684	1.447	0.018	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc	→	GENERO=hombre
0.052	0.889	0.058	8.611	1.778	0.023	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales	→	TIPOS_ESPACIOS=Oficinas
0.048	0.882	0.055	9.118	1.765	0.021	CARRERA=Ingeniería en Computación, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Sin Carrera	→	TIPOS_ESPACIOS=Oficinas
0.048	0.938	0.052	12.688	1.432	0.015	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.045	1.000	0.045	11.071	2.000	0.023	GENERO=mujer, CARRERA=Ingeniería en Computación, GRADO_DE_PASION=REGULAR	→	TIPOS_ESPACIOS=Oficinas
0.045	0.875	0.052	9.688	1.750	0.019	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Ninguna, GRADO_DE_PASION=MUCHO	→	TIPOS_ESPACIOS=Oficinas
0.045	0.875	0.052	12.688	1.336	0.011	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, CARR_PROF_PADRES=S/carrera	→	GENERO=hombre
0.045	0.933	0.048	13.533	1.425	0.013	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→	GENERO=hombre
0.045	0.875	0.052	12.688	1.336	0.011	INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.042	1.000	0.042	15.615	1.527	0.014	INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=REGULAR	→	GENERO=hombre
0.042	0.867	0.048	13.533	1.323	0.010	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.042	0.867	0.048	8.800	2.035	0.021	GENERO=hombre, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Sociales	→	CARRERA=Ingeniería en Computación
0.042	0.867	0.048	10.333	1.733	0.018	GENERO=hombre, CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales	→	TIPOS_ESPACIOS=Oficinas
0.039	0.857	0.045	11.071	1.714	0.016	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Ninguna, CARR_PROF_PADRES=Sin Carrera	→	TIPOS_ESPACIOS=Oficinas
0.039	0.857	0.045	11.071	1.714	0.016	GENERO=hombre, CARRERA=Ingeniería en Computación, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Sin Carrera	→	TIPOS_ESPACIOS=Oficinas
0.039	0.923	0.042	15.615	1.410	0.011	TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→	GENERO=hombre
0.039	0.923	0.042	15.615	1.410	0.011	INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=S/carrera	→	GENERO=hombre
0.035	0.917	0.039	16.917	1.400	0.010	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.035	0.917	0.039	16.917	1.400	0.010	TIPOS_ESPACIOS=Otros	→	GENERO=hombre
0.032	0.909	0.035	9.818	2.609	0.020	GENERO=mujer, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=S/carrera	→	INFLU_MEDIATICA=Ninguno
0.032	0.909	0.035	15.727	1.629	0.012	GENERO=mujer, INFLU_MEDIATICA=Ninguna, CARR_PROF_PADRES=Sin Carrera	→	GRADO_DE_PASION=MUCHO
0.032	0.909	0.035	18.455	1.388	0.009	CARRERA=Ingeniería en Computación, TIPOS_ESPACIOS=Oficinas, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO	→	GENERO=hombre
0.032	0.909	0.035	18.455	1.388	0.009	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, GRADO_DE_PASION=MUCHO, CARR_PROF_PADRES=Diferente	→	GENERO=hombre
0.032	1.000	0.032	20.300	1.527	0.011	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Soc, CARR_PROF_PADRES=S/carrera	→	GENERO=hombre
0.032	0.909	0.035	14.091	1.818	0.015	CARRERA=Ingeniería en Computación, INFLU_MEDIATICA=Web / Redes Sociales, CARR_PROF_PADRES=Diferente	→	TIPOS_ESPACIOS=Oficinas

Con el algoritmo de reglas de asociación, se identificaron reglas con soporte mínimo de 3% y confianza mínima del 85% con los atributos categóricos del conjunto de datos. Se hallaron un total de 30 reglas. De forma interesante, de las reglas resultantes se encontró que la carrera de Ingeniería en Computación está fuertemente asociada al género masculino. Otra regla interesante identificada fue que las mujeres que tomaron la decisión de carrera sin influencia mediática tienen un alto grado de pasión por sus estudios. Otra regla más evidente, es que se asocia la carrera de ingeniería en computación con un tipo de espacio cerrado (una oficina).

5. Identificación de Clases

	CN2 rule inducer	error	Naive Bayes	error
291	<u>0.17 : 0.07 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.04 : 0.30</u> → MALO	<u>0.330</u>
292	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.63 : 0.37</u> → MALO	<u>0.374</u>
293	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.47 : 0.53</u> → REGU...	<u>0.469</u>
294	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.47 : 0.00 : 0.53</u> → REGU...	<u>0.533</u>
295	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.62 : 0.38</u> → MALO	<u>0.618</u>
296	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.37 : 0.63</u> → REGU...	<u>0.633</u>
297	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.21 : 0.00 : 0.79</u> → REGU...	<u>0.788</u>
298	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.81 : 0.19</u> → MALO	<u>0.807</u>
299	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.83 : 0.17</u> → MALO	<u>0.833</u>
300	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.87 : 0.13</u> → MALO	<u>0.868</u>
301	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.96 : 0.04</u> → MALO	<u>0.958</u>
302	<u>0.17 : 0.67 : 0.17</u> → MALO	<u>0.333</u>	<u>0.00 : 0.96 : 0.04</u> → MALO	<u>0.962</u>
303	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.00 : 0.97 : 0.03</u> → MALO	<u>0.974</u>
304	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.02 : 0.00 : 0.97</u> → REGU...	<u>0.977</u>
305	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.02 : 0.00 : 0.98</u> → REGU...	<u>0.978</u>
306	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.02 : 0.00 : 0.98</u> → REGU...	<u>0.983</u>
307	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.01 : 0.00 : 0.99</u> → REGU...	<u>0.994</u>
308	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.01 : 0.00 : 0.99</u> → REGU...	<u>0.994</u>
309	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.00 : 0.00 : 1.00</u> → REGU...	<u>0.995</u>
310	<u>0.20 : 0.60 : 0.20</u> → MALO	<u>0.400</u>	<u>0.00 : 0.00 : 1.00</u> → REGU...	<u>0.997</u>
311	<u>0.01 : 0.01 : 0.98</u> → REGU...		<u>1.00 : 0.00 : 0.00</u> → BUENO	

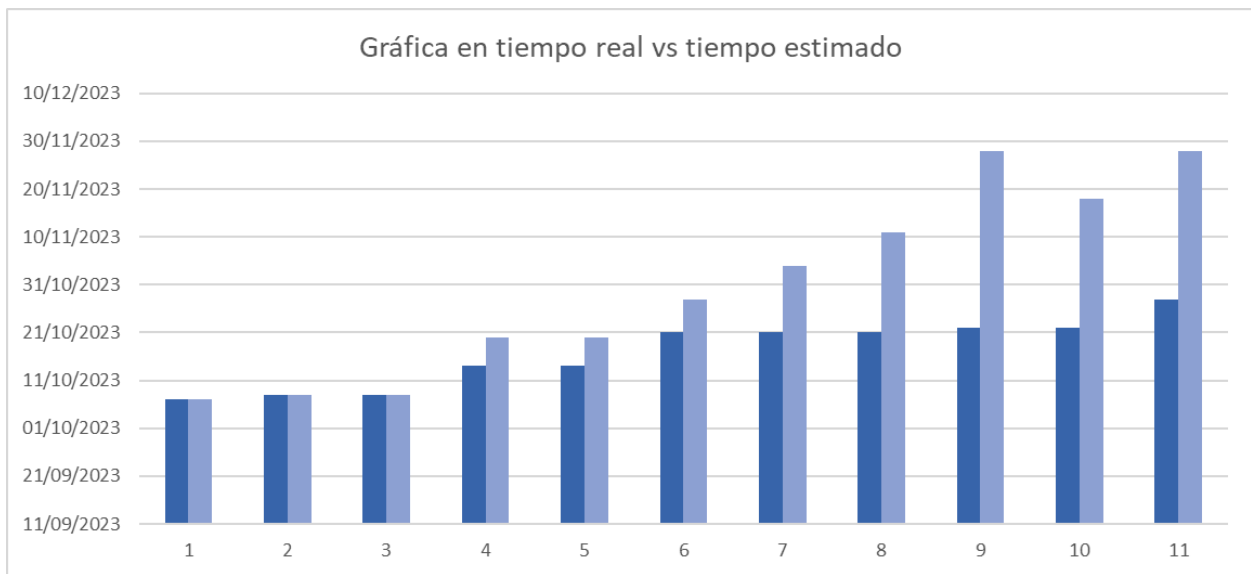
Para predecir la clasificación del estudiante en esta caso podemos observar que el CN2 genera porcentajes de errores menores caso contrario de Naive Bayes. Por lo que si tuviéramos que predecir la clasificación de nuevos estudiantes lo mejor sería CN2.

Ejecutando el plan de trabajo: fecha planeada vs fecha real

El plan se ejecutó tratando de establecer los plazos comprometidos, sin embargo, hubo circunstancias que ocasionaron retrasos. Como resultado, se concluyó en el tiempo de entrega comprometido pero con una variación de fechas. A continuación, se presenta el plan de trabajo.

No	Actividad	Fecha Estimada	Fecha Real	Tiempo estimado	Diferencia de días
1	Elección del proyecto	07 de octubre 2023	07 de octubre 2023	1 semana	Inició en tiempo
2	Planificación	08 de octubre 2023	08 de octubre 2023	1 semana	Inició en tiempo
3	Creación del documento	08 de octubre 2023	08 de octubre 2023	1 semana	Inició en tiempo
4	Problema de Negocio	14 de octubre 2023	20 de octubre 2023	2 semanas	Atraso de 6 días
5	Elementos de información	14 de octubre 2023	20 de octubre 2023	2 semanas	Atraso de 6 días
6	Estadísticos	21 de octubre 2023	28 de octubre 2023	1 semana	Atraso de 7 días
7	Métodos de limpieza	21 de octubre 2023	04 de noviembre 2023	2 semanas	Atraso de 14 días
8	Aplicación de modelos	21 de octubre 2023	11 de noviembre 2023	3 semanas	Atraso de 21 días
9	Documentar	22 de octubre 2023	28 de noviembre 2023	4 semanas	Atraso de 30 días
10	Identificación de clases	22 de octubre 2023	18 de noviembre 2023	3 semanas	Atraso de 27 días
11	Presentación	28 de octubre 2023	28 de noviembre 2023	4 semanas	Atraso de 30 días

Plan de trabajo



Gráfica tiempo real vs tiempo estimado

Como podemos observar, se definió un plan de fechas estimado, sin embargo, vemos que tenemos diferencias de días notables, esto puede sonar preocupante por el tema de SLA's vencidos de acuerdo a lo que se había definido, sin embargo, al tener una fecha compromiso prevista para la entrega del proyecto (29 de noviembre), se logró concluir en la fecha marcada.

Recordemos que dentro de nuestro proyecto se ve afectado por varias circunstancias, entre ellas se cruzan días feriados, proyectos de otras materias o cargas de trabajo adicionales, suspensión de actividades o actividades extracurriculares. Pese a ello se respetó el deadline de la entrega del proyecto conforme a lo acordado.

Propuesta de negocio

La presente propuesta de negocio, tiene como finalidad ofrecer un programa integral de desarrollo académico para estudiantes de nuevo ingreso o aquellos que se encuentren iniciando una carrera universitaria, la iniciativa está basada en la identificación temprana de las necesidades y cuestiones de los estudiantes, haciendo uso de los datos obtenidos a través de una encuesta de elección de carrera, el proyecto se desarrollará en colaboración con instituciones educativas para brindar cursos específicos que fortalezcan los conocimientos antecedentes, habilidades técnicas y blandas, permitiendo un mejor desempeño en los primeros semestres y en toda la trayectoria universitaria.

Como objetivos se tendrán identificar a los estudiantes de nuevo ingreso o próximos a ingresar para ser beneficiados con cursos de materias esenciales y habilidades específicas, ofrecer cursos propedéuticos para encarar áreas de oportunidad, utilizar los datos recopilados para orientar y enriquecer los programas de desarrollo académico.

Nuestra metodología:

- **Detección anticipada**
 - Identificar a instituciones educativas que puedan beneficiarse con el programa
 - Enriquecer los datos que se tienen para personalizar la oferta de cursos
- **Desarrollo de contenidos**
 - Crear y diseñar cursos específicos basados en los resultados de la encuesta y necesidades identificadas
 - Incluir enfoque pedagógico que contribuya el aprendizaje efectivo
- **Llevar a la práctica los cursos:**
 - Impartir cursos propedéuticos antes del inicio del semestre como preparación a los estudiantes
 - Mantener una constante evaluación del progreso para realizar ajustes a las necesidades emergentes.
- **Supervisión y evaluación:**
 - Realizar encuestas de opinión de los estudiantes para implementar mejoras continuas en la oferta de cursos
 - Realizar supervisión durante los primeros semestres para evaluar el impacto de los cursos en el rendimiento académico.

Beneficios:

- **Progreso académico:**
 - Equipar a los estudiantes con herramientas y conocimientos que mejoren su rendimiento desde el inicio de su trayectoria académica.
- **Toma de decisiones informada:**
 - Adaptar la oferta de cursos y brindar orientación personalizada mediante el análisis de datos recopilados en encuestas.
- **Asociación con instituciones educativas:**
 - Fomentar una colaboración sólida con instituciones educativas para respaldar sus esfuerzos en el desarrollo estudiantil.
- **Apoyo a la toma de decisiones profesionales:**
 - Ayudar en la toma de decisiones informadas sobre elección de carreras profesionales al proporcionar una exposición temprana a las dinámicas del mercado laboral.

Business Case

Este proyecto resulta extremadamente intrigante, ya que nos brinda la oportunidad de comprender los elementos que inciden en la toma de una decisión crucial en la vida: la elección de una carrera universitaria. Diversos factores, como los consejos de familiares y amigos, el desempeño académico en la preparatoria, el salario promedio asociado a la carrera, la distancia hasta la universidad, el costo de los estudios, entre otros, pueden influir en esta decisión. Estos elementos determinarán nuestro nivel de satisfacción y, a su vez, nuestra posición económica en el futuro. Por ende, este proyecto será de gran utilidad al permitir que las personas identifiquen su camino, tomen decisiones más informadas y utilicen los factores relevantes a su favor al seleccionar su carrera.

El modelo de negocio de este proyecto implica transformar la información recopilada en una aplicación de pago único (suscripción), lo que facilitaría a los estudiantes indecisos encontrar la orientación adecuada para elegir su carrera. Otra posibilidad de modelo sería la comercialización del proyecto con diversas instituciones educativas, permitiéndoles ofrecerlo a sus propios alumnos.

A continuación se muestran los costos que tendría nuestro proyecto de acuerdo a nuestro plan de trabajo enfocándonos en el entorno técnico como lo son los estadísticos, limpieza de datos, aplicación de modelos, etc.

En la imagen que se muestra debajo, se puede observar que el costo estimado del proyecto asciende a \$113,401.60 pesos. Esto se debe a la necesidad de contratar 4 mineros de datos Jr y dedicar un total de 200 horas al proyecto.

No	Descripción	Cantidad
1	Personas que intervinieron en el proyecto	4 personas (Jrs)
2	Tiempo invertido en el proyecto	1.2 semanas (8 días hábiles)
3	Tiempo invertido en el proyecto por horas	200 horas
4	Sueldo promedio de un minero de datos por mes	\$15,000.00
5	Sueldo promedio de un minero de datos por día	\$500.00
6	Sueldo promedio de un minero de datos por hora	\$62.50

No	Descripción	Cantidad
1	Cantidad invertida por personas (horas*pesos)	\$12,500.00
2	Cantidad invertida por 4 personas	\$50,000.00
3	Costo de software por 4 personas	\$3,200.00
4	Consumo de servicios (luz, internet, etc)	\$4,000.00
5	Costo de la base de datos	\$5,500.00
6	Iva	\$12,032.00
7	Total Invertido	\$87,232.00
8	Total Proyecto con ganancia del 30%	\$113,401.60

Se realizó además un total invertido del proyecto y un total con la ganancia al 30% esto para tener una holgura respecto a lo que se va a pagar vs lo que se va a recuperar.

Desde una perspectiva financiera, el enfoque de vender el modelo a las escuelas no genera ganancias directas. No obstante, su valor reside en el respaldo que brinda a los estudiantes a través de un conjunto de algoritmos y bases de datos analizadas, mejorando así su capacidad para tomar decisiones el cual es un aspecto crucial en sus vidas.

Conclusión

El proyecto presentado se desarrolló con un enfoque orientado a los estudiantes en su desarrollo académico que va más allá de la enseñanza tradicional, abordando sus necesidades particulares desde el inicio de su carrera universitaria; la pieza clave de nuestra estrategia radica en una colaboración estrecha y activa con instituciones educativas, uniendo fuerzas para moldear un entorno educativo más receptivo y personalizado. Al utilizar datos significativos recopilados mediante encuestas y análisis detallados, nuestra aspiración es crear programas y apoyos que no solo satisfagan las demandas académicas, sino que también aborden obstáculos que podrían conducir a la deserción escolar, así como cultivar una conexión apasionada entre los estudiantes y sus elecciones de carrera.

La pasión es el motor que impulsa la dedicación y el compromiso duradero, mediante intervenciones estratégicas y programas de apoyo, aspiramos a difundir esta pasión, inspirando a los estudiantes a no solo completar sus estudios, sino a encontrar una vocación que verdaderamente los apasione y les brinde satisfacción en su camino profesional.