



ACTIVIDAD GUIADA 3

CRISP-DM

Máster en Big Data y Data Science

13MBID – Metodologías de Gestión y Diseño de
Proyectos Big Data

Nombre: Ingrid Julieth Rodriguez Nova

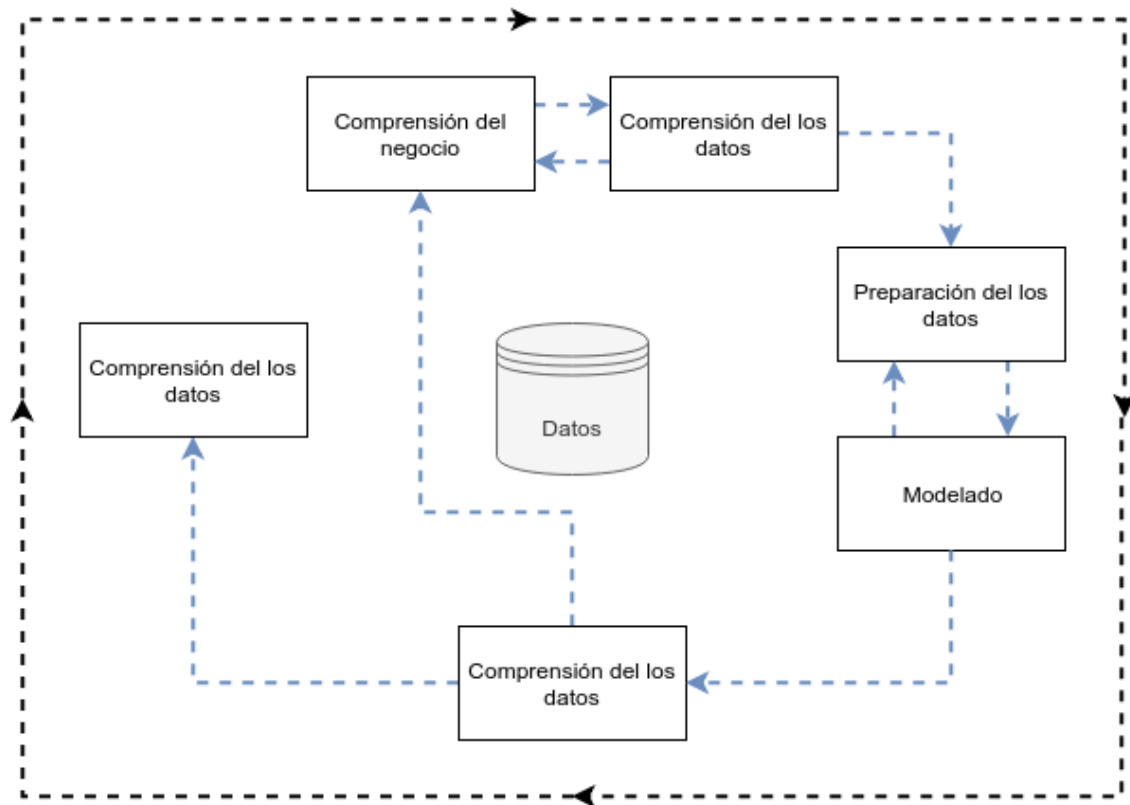
Fecha: Julio 14 del 2023

Seminario II - Aplicando técnicas ágiles para la gestión de proyectos de ciencia de datos

El presente documento es una planilla que se utilizará para el desarrollo de la documentación correspondiente a las actividades del Seminario II y su correspondiente Actividad Guiada. El contenido será guiado según las fases y actividades de la metodología CRISP-DM.

Una vez completado con la información correspondiente al proyecto de ciencia de datos y complementado con los reportes de la ejecución de la libreta Jupyter desarrollada se podrán finalizar las tareas del proyecto.

La metodología CRISP-DM cuenta con 6 fases, ver figura 1, que forman un ciclo iterativo, con vistas a lo que se podrá considerar como un proceso iterativo-incremental de desarrollo de soluciones de ciencia de datos para un contexto en particular.



[A] Fase de comprensión del problema

- **Determinar los objetivos de la Organización**

Las autoridades de una Institución Universitaria desean obtener conocimiento a partir de los datos disponibles de los alumnos, principalmente en lo que respecta a su situación como estudiante **Activo** (continúa cursando la carrera) / **Pasivo** (ha abandonado los estudios o al menos no se ha reinscrito para continuar cursando en la actualidad). El objetivo final que se persigue es el de poder predecir con un margen de confianza considerable la situación de los nuevos alumnos inscritos para el periodo 2022.

- **Evaluación de la situación**

Se cuenta con los siguientes recursos para la ejecución del proyecto:

- Los datos históricos de los estudiantes distribuidos en las tres dimensiones de análisis (inscripción, cursado del 1er. semestre y datos académicos al final el año)
- Se cuenta con el personal para la realización de las tareas del proyecto.
- Se cuenta con las herramientas y equipos necesarios para la experimentación e implementación requerida.

- **Determinación de los objetivos del proyecto**

Elaborar un **modelo de predicción** de la situación académica de los alumnos de la Unidad Académica (Facultad / Escuela) en cuestión en base a los datos históricos de la misma.

La efectividad mínima del modelo de predicción deberá ser del **80%**.

- **Definir plan del proyecto**

Se definen las fases de la metodología CRISP-DM como épicas (Epic) y las tareas como ítems de la pila de productos (Product Backlog Item (PBI)), se distribuyen en 2 Sprint cada uno con tres fases y se les asigna un responsable del equipo de trabajo, para el ejemplo se asigna a “Ingrid Rodriguez Nova”.

13MBID-Metodología CRISP-DM Team							
Backlog Analytics + New Work Item View as Board Column Options ...							
Order	Work Item T...	Title	State	Effort	Assigned To	Iteration Path	
1	Epic	Compreñsion del negocio	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Determinar los objetivos de la organización	Done	1	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Evaluación de la situación	Done	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Determinación de los objetivos del proyecto	Done	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Definir plan del proyecto	Approv...	8	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
2	Epic	Compreñsion de los datos	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Recolección de datos iniciales	New	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Descripción de los datos	New	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Exploración de datos	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Verificación de la calidad de los datos	New	8	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
3	Epic	Preparación de los datos	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Selección de datos	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Limpieza de los datos	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Construcción de los datos	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Integración de los datos	New	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	
	Product Ba...	Formateo de los datos	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 1	

Imagen 1 - Planeación Sprint 1

13MBID-Metodología CRISP-DM Team							
Backlog Analytics + New Work Item View as Board Column Options ...							
Order	Work Item T...	Title	State	Effort	Assigned To	Iteration Path	
4	Epic	Modelado	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Selección de la técnica de modelado	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Generar el diseño de las pruebas	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Construcción del modelo	New	8	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Evaluar el modelo	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
5	Epic	Evaluación	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Evaluar los resultados	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Revisar el proceso	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Determinar pasos siguientes	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
6	Epic	Despliegue	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Planificar el despliegue	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Planificar monitoreo y mantenimiento	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Producir el reporte final	New	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Revisar el proyecto	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	

Imagen 2 - Planeación Sprint 2

Para cumplir con el objetivo de cada PBI se establecen una serie de tareas. A continuación, se pueden ver las tareas (Task) de los PBIs del Sprint 1:

13MBID-Metodología CRISP-DM Team ▾ ☆ 🔍						
Backlog Analytics + New Work Item 🔗 View as Board 🔗 Column Options ...						
Order	Work Item T...	Title	State	Effort	Assigned To	
1	Epic	Comprensión del negocio	New		Ingrid Rodriguez Nova	
+	Product Ba...	Determinar los objetivos de la organiz...	Done	1	Ingrid Rodriguez Nova	
	Task	Revisar los objetivos del negocio	Done		Ingrid Rodriguez Nova	
	Product Ba...	Evaluación de la situación	Done	2	Ingrid Rodriguez Nova	
	Task	Reconocer los recursos de información ...	Done		Ingrid Rodriguez Nova	
	Product Ba...	Determinación de los objetivos del proyecto	Done	3	Ingrid Rodriguez Nova	
	Task	Especificar los objetivos de efectividad d...	Done		Ingrid Rodriguez Nova	
	Task	Especificar los resultados esperados a o...	Done		Ingrid Rodriguez Nova	
	Product Ba...	Definir plan del proyecto	Approv...	8	Ingrid Rodriguez Nova	
	Task	Especificar tareas	In Progr...		Ingrid Rodriguez Nova	
	Task	Especificar tiempos	In Progr...		Ingrid Rodriguez Nova	
	Task	Especificar recursos humanos y materiales	In Progr...		Ingrid Rodriguez Nova	

Imagen 3 – Tareas Sprint 1 – Fase 1 Comprensión del negocio

13MBID-Metodología CRISP-DM Team ▾ ☆ 🔍						
Backlog Analytics + New Work Item 🔗 View as Board 🔗 Column Options ...						
Order	Work Item T...	Title	State	Effort	Assigned To	
2	Epic	Comprensión de los datos	New		Ingrid Rodriguez Nova	
	Product Ba...	Recolección de datos iniciales	New	2	Ingrid Rodriguez Nova	
	Task	Recuperar los archivos de datos	To Do		Ingrid Rodriguez Nova	
	Task	Verificar la integridad de los archivos	To Do		Ingrid Rodriguez Nova	
	Product Ba...	Descripción de los datos	New	2	Ingrid Rodriguez Nova	
	Task	Detallar composición a nivel de atributo...	To Do		Ingrid Rodriguez Nova	
	Task	Describir número de filas	To Do		Ingrid Rodriguez Nova	
	Product Ba...	Exploración de datos	New	3	Ingrid Rodriguez Nova	
	Task	Descripción de metadatos por dataset	To Do		Ingrid Rodriguez Nova	
	Product Ba...	Verificación de la calidad de los datos	New	8	Ingrid Rodriguez Nova	
	Task	Verificar valores nulos o faltantes	To Do		Ingrid Rodriguez Nova	
	Task	Verificar cumplimiento de formatos de d...	To Do		Ingrid Rodriguez Nova	
	Task	Verificar rangos de valores	To Do		Ingrid Rodriguez Nova	

Imagen 4 – Tareas Sprint 1 – Fase 2 Comprensión de los datos

Order	Work Item T...	Title	State	Effort	Assigned To
3	Epic	Preparación de los datos	New		Ingrid Rodriguez Nova
	Product Ba...	Selección de datos	New	3	Ingrid Rodriguez Nova
	Task	Determinar las columnas a eliminar ...	To Do		Ingrid Rodriguez Nova
	Product Ba...	Limpieza de los datos	New	3	Ingrid Rodriguez Nova
	Task	Aplicar filtros para limpieza a nivel de filas	To Do		Ingrid Rodriguez Nova
	Product Ba...	Construcción de los datos	New	5	Ingrid Rodriguez Nova
	Task	Determinar los atributos nuevos que se ...	To Do		Ingrid Rodriguez Nova
	Product Ba...	Integración de los datos	New	2	Ingrid Rodriguez Nova
	Task	Generar dataset unificado	To Do		Ingrid Rodriguez Nova
	Product Ba...	Formateo de los datos	New	3	Ingrid Rodriguez Nova
	Task	Definir y aplicar adaptaciones en los datos...	To Do		Ingrid Rodriguez Nova

Imagen 5 – Tareas Sprint 1 – Fase 3 Preparación de los datos

En las imágenes anteriores se puede evidenciar que las tareas están asignadas a “Ingrid Rodriguez Nova”, en la siguiente tabla se describe los roles que desempeñara durante el proyecto:

Rol	Descripción de Responsabilidades	Persona
Científico de datos	<ul style="list-style-type: none"> ○ Manipulación y explotación datos ○ Uso de herramientas para machine learning, estadística, matemáticas entre otros. 	Ingrid Rodriguez Nova
Ingeniero / Arquitecto de datos:	<ul style="list-style-type: none"> ○ Se encarga de preparar los datos para que los usuarios finales puedan hacer uso de ellos. ○ Tiene un gran conocimiento de los datos y normalmente como se relacionan con el negocio. ○ Participa activamente en el contexto de Gobierno de Datos 	Ingrid Rodriguez Nova
Analista de negocios:	<ul style="list-style-type: none"> ○ Tiene un gran conocimiento del negocio y sirve de intermediario entre el equipo técnico y los usuarios. ○ Conocimientos de funcionalidades y responsabilidades de las herramientas tecnológicas. 	Ingrid Rodriguez Nova

Gestores de proyectos:	<ul style="list-style-type: none"> ○ Deben asegurar la ejecución de las tareas planteadas en el plan de proyecto para así dar solución al problema planteado. ○ Debe trabajar en pro de optimizar los recursos de los que dispone el proyecto. 	Ingrid Rodriguez Nova
------------------------	--	-----------------------

Tabla 1 - Roles y personas del proyecto

La descripción de las responsabilidades anteriormente descritas está basada en el Anexo “Equipo de Ciencia de Datos” proporcionados por los docentes de la materia. Aunque para el efecto de esta practica se relaciono la misma persona en los 4 roles, se recomienda para proyectos de ejecución real que de preferencia el Gestor de Proyectos sea una persona especializada en este rol al igual que el analista de negocios. Para el caso de los otros dos roles (Científico de datos e Ingeniero de Datos) si es un proyecto medianamente pequeño la misma persona puede desempeñar ambos roles.

[B] Fase de comprensión de los datos

- **Recolección de datos iniciales**

Se han recuperado los datos históricos agrupados en 3 dimensiones / archivos:

- **Datos de inscripciones:** referidos a la inscripción de cada estudiante a alguna de las carreras de la unidad académica.
- **Datos del 1er cuatrimestre:** referidos al rendimiento académico de cada estudiante en su primer cuatrimestre de cursado y el primer periodo de exámenes.
- **Datos académicos:** referidos al avance en la carrera de cada estudiante y su situación académica actualizada

Los datos han sido verificados contra los orígenes y se encuentran aptos para su uso.

- **Descripción de los datos**

Se describen las características principales de cada dataset:

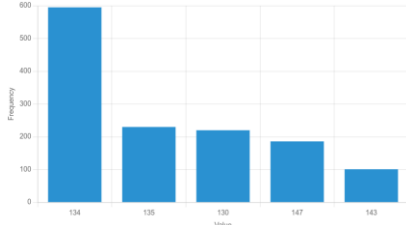
Dataset	Columnas / Atributos	Cantidad de filas
datos_inscripciones	id_estudiante propuesta estado_inscripcion plan_estudios version_plan modalidad fecha_ingreso fecha_inscripcion	Cantidad de filas: 1332
datos_cursado	id_estudiante estado_inscripcion ingreso_aprobadas ingreso_libres ingreso_totales cursadas_aprobadas cursadas_regulares cursadas_libres cursadas_totales inscripciones_exámenes exámenes_aprobados	Cantidad de filas: 1332
datos_academicos	id_estudiante plan anio_ingreso fecha_ingreso fecha_ultimo_examen	Cantidad de filas: 815

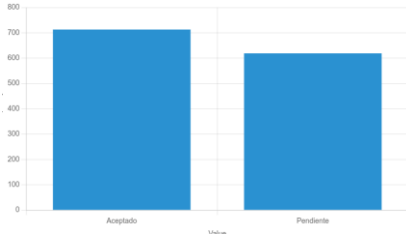
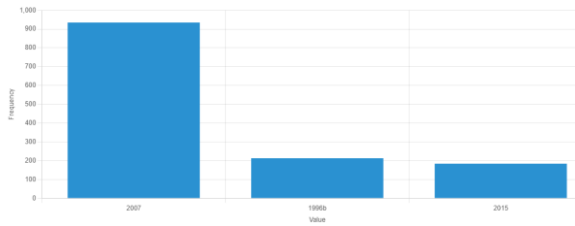
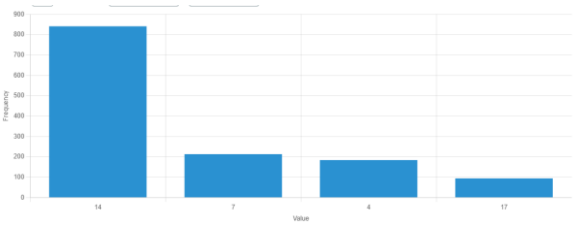
	anio_ultima_reinscripcion promedio_sin_aplazos promedio_con_aplazos actividades_aprobadas total_actividades regular calidad segundo_anio	
--	---	--

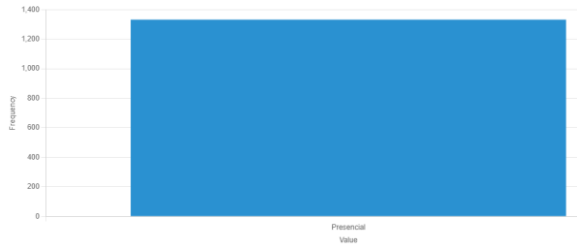
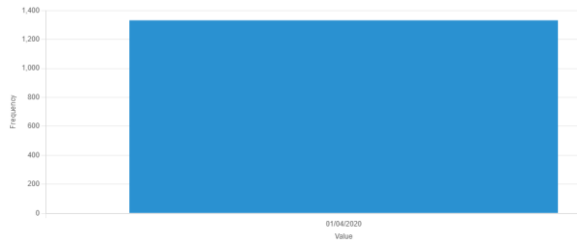
- **Exploración de datos**

Se describen a continuación los meta-datos de cada dataset:

- **datos_inscripciones**

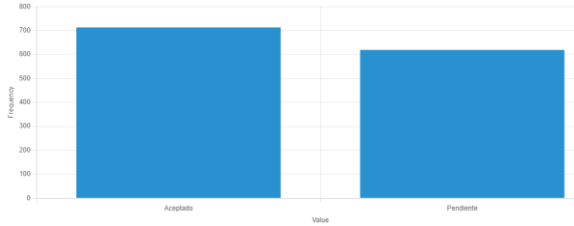
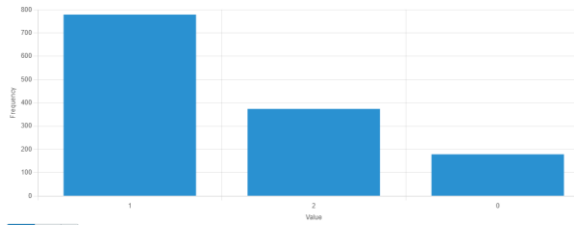
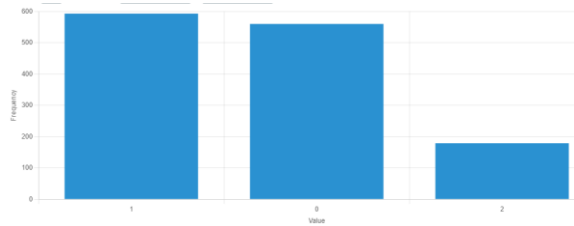
Columna / Atributo	Tipo de datos	Observaciones
id_estudiante	String / Alfanumérico	<p>Atributo con un formato especial: ##-#####-#</p> <p>Valores de ejemplo: CA-000281-5, CA-003269-2, CA-004968-4, CA-005300-2, CA-006018-1</p> <p>Hace referencia al número de matrícula de cada estudiante.</p> <p>Cantidad de nulos = 0</p>
propuesta	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>134 (595) 135 (230) 130 (220) 147 (186) 143 (101)</p>  <p>Cantidad de nulos = 0</p>
estado_inscripcion	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p>

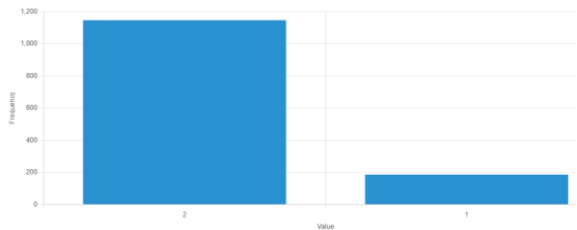
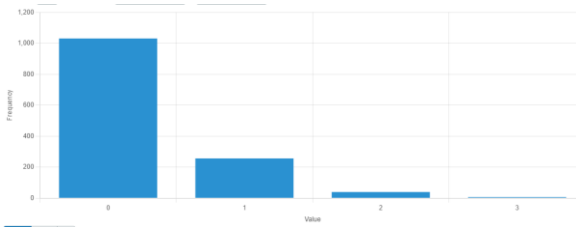
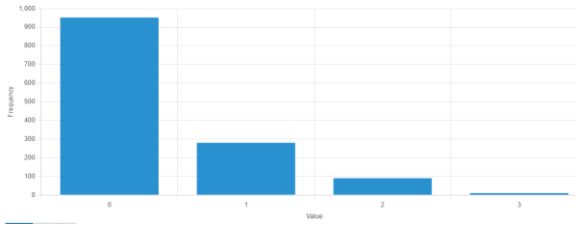
		<p>Aceptado (713) Pendiente (619)</p>  <p>Cantidad de nulos = 0</p>
plan_estudios	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>2007 (935) 1996b (213) 2015 (184)</p>  <p>Cantidad de nulos = 0</p>
version_plan	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>14 (841) 7 (213) 4 (184) 17 (94)</p>  <p>Cantidad de nulos = 0</p>
modalidad	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>Presencial (1332)</p>

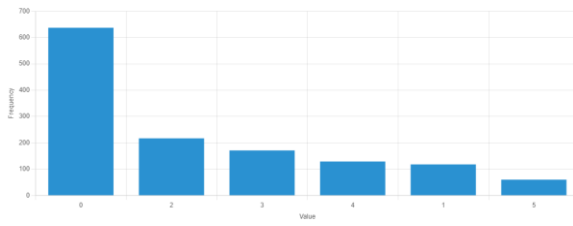
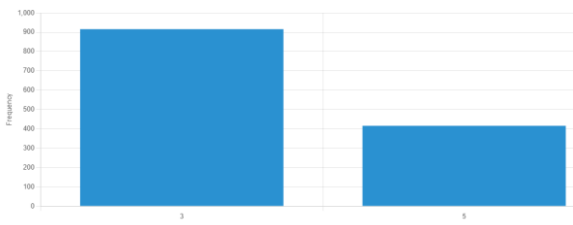
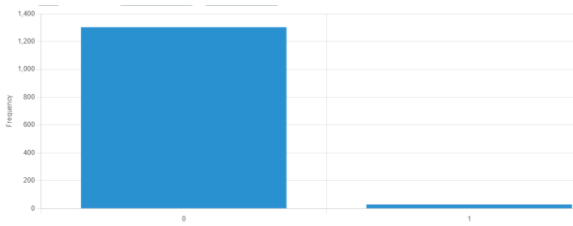
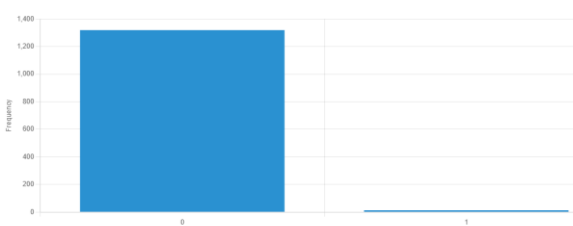
		 <p>Cantidad de nulos = 0</p>
fecha_ingreso	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>01/04/2020 (1332)</p>  <p>Formato de fecha: <i>dd/mm/aaaa</i> Cantidad de nulos = 0</p>
fecha_inscripcion	String / Alfanumérico	<p>Formato de fecha: <i>dd/mm/aaaa</i></p> <p>La cantidad de valores diferentes es: 77</p> <p>Los más frecuentes son: 04/02/2020 (73) 14/02/2020 (71) 18/12/2019 (68) 17/12/2019 (65)</p> <p>Cantidad de nulos = 0</p>

▪ **datos_cursado**

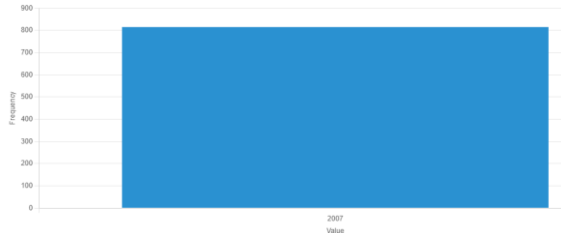
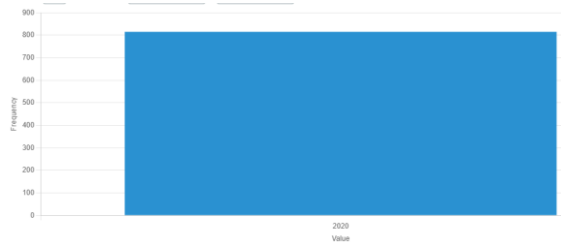
Columna / Atributo	Tipo de datos	Observaciones
id_estudiante	String / Alfanumérico	<p>Atributo con un formato especial: ##-#####-#</p> <p>Valores de ejemplo: CA-000281-5 (1), CA-003269-2 (1), CA-004968-4 (1), CA-005300-2 (1), CA-006018-1 (1)</p> <p>Hace referencia al número de matrícula de cada</p>


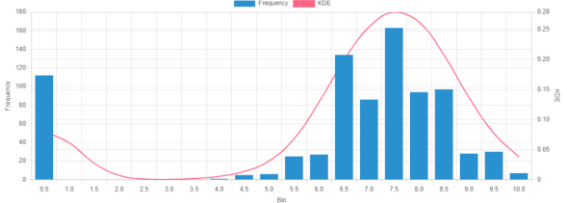
		<p>estudiante.</p> <p>Es un valor único en la tabla.</p> <p>Cantidad de nulos = 0</p>
estado_inscripcion	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>Aceptado (713) Pendiente (619)</p>  <p>Cantidad de nulos = 0</p>
ingreso_aprobadas	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>1 (779) 2 (374) 0 (179)</p>  <p>Cantidad de nulos = 0</p>
ingreso_libres	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>1 (593) 0 (560) 2 (179)</p>  <p>Cantidad de nulos = 0</p>
ingreso_totales	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p>

		<p>2 (1146) 1 (186)</p>  <p>Cantidad de nulos = 0</p>
cursadas_aprobadas	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (1031) 1 (256) 2 (39) 3 (6)</p>  <p>Cantidad de nulos = 0</p>
cursadas_regulares	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (952) 1 (280) 2 (90) 3 (10)</p>  <p>Cantidad de nulos = 0</p>
cursadas_libres	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (637) 2 (217) 3 (171) 4 (129) 1 (118) 5 (60)</p>

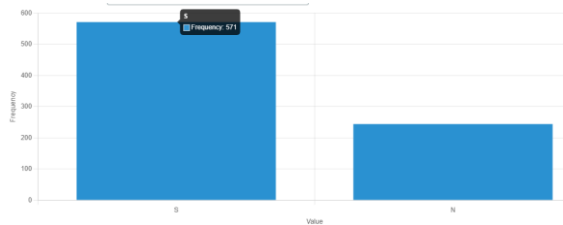
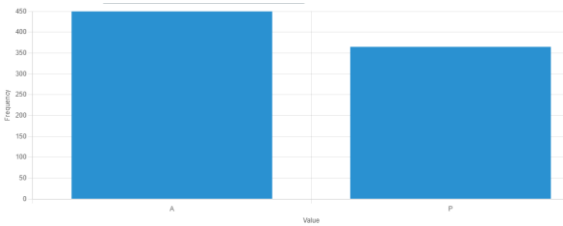
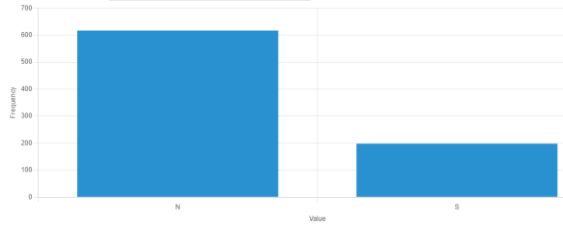
		 <p>Cantidad de nulos = 0</p>
cursadas_totales	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>3 (916) 5 (416)</p>  <p>Cantidad de nulos = 0</p>
inscripciones_examenes	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (1303) 1 (29)</p>  <p>Cantidad de nulos = 0</p>
exámenes_aprobados	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (1319) 1 (13)</p>  <p>Cantidad de nulos = 0</p>

▪ **datos_academicos**

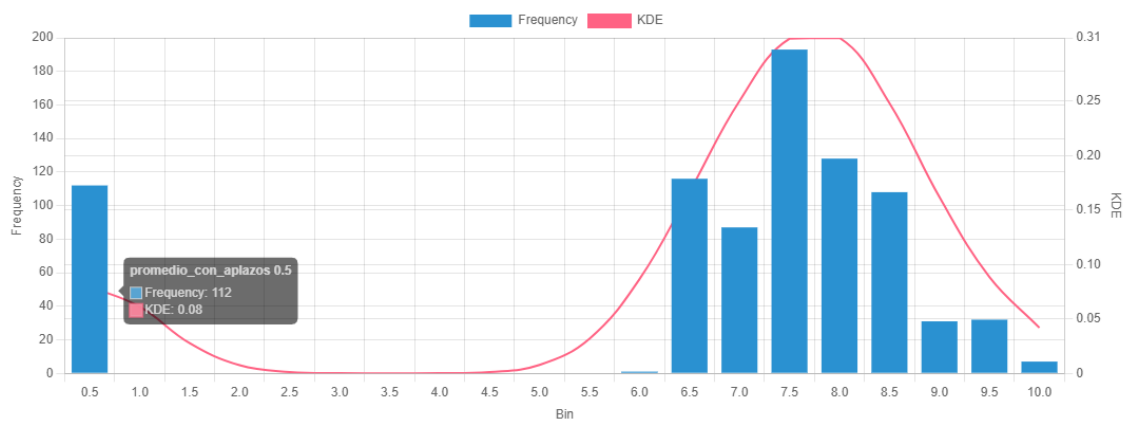
Columna / Atributo	Tipo de datos	Observaciones
id_estudiante	String / Alfanumérico	<p>Atributo con un formato especial: ##-#####-#</p> <p>Valores de ejemplo: CA-000281-5 (1), CA-003269-2 (1), CA-004968-4 (1), CA-005300-2 (1), CA-006018-1 (1)</p> <p>Hace referencia al número de matrícula de cada estudiante.</p> <p>Es un valor único en la tabla.</p> <p>Cantidad de nulos = 0</p>
plan	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>2007 (815)</p>  <p>Cantidad de nulos = 0</p>
anio_ingreso	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>2020 (815)</p>  <p>Cantidad de nulos = 0</p>
fecha_ingreso	String /	Descripción y distribución de valores

	Alfanumérico	<p>(cantidad):</p> <p>Formato de fecha: <i>dd/mm/aaaa</i></p> <p>01/04/2020 (815)</p>  <p>Cantidad de nulos = 0</p>
fecha_ultimo_examen	String / Alfanumérico	<p>La cantidad de valores diferentes es: 24</p> <p>Los más frecuentes son:</p> <p>18/08/2021 (48) 20/08/2021 (40) 23/08/2021 (36) 19/08/2021 (16) 16/03/2021 (12)</p> <p>Cantidad de nulos = 596 (73.13%) Datos con valor = 219 (26.87%)</p>
anio_ultima_reinscripcion	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>2021.0 (450)</p> <p>Se valida con el experto del dominio y el tipo de dato está mal, debe ser un entero porque hace referencia al año de la última inscripción de un estudiante.</p> <p>Cantidad de nulos = 365 (44.79%) Datos con valor = 450 (55.21%)</p>
promedio_sin_aplazos	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (10)</p> 

		<p>La mayoría de las notas de los estudiantes entre 6.5 y 8.5</p> <p>Es probable que los 112 estudiantes con promedio cerca del cero (0) no se matriculen al segundo año.</p> <p>Cantidad de nulos = 0</p>
promedio_con_aplazos	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (10)</p>  <p>La mayoría de las notas de los estudiantes entre 6.5 y 8.5</p> <p>Cantidad de nulos = 0</p>
actividades_aprobadas	Integer / Numérico	<p>La cantidad de valores diferentes es: 30</p> <p>Los más frecuentes son (cantidad):</p> <p>2 (217)</p> <p>1 (132)</p> <p>0 (112)</p> <p>3 (99)</p> <p>4 (57)</p> <p>Cantidad de nulos = 0</p> <p>Datos con valor = 815</p>
total_actividades	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>40 (595)</p> <p>42 (220)</p>  <p>Cantidad de nulos = 0</p>

regular	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>S (571) N (244)</p>  <p>Cantidad de nulos = 0</p>
calidad	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>A (450) P (365)</p>  <p>Cantidad de nulos = 0</p>
segundo_anio	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>A (450) P (365)</p>  <p>Cantidad de nulos = 0</p>

De la exploración anterior de los datos se quiere resaltar el promedio de los estudiantes porque se evidencia a través del siguiente histograma que los 112 estudiantes con promedio cero (0) probablemente sean desertores para el segundo año.



- **Verificación de la calidad de los datos**

Se procede a verificar la calidad de los datos de los tres datasets. Los resultados se describen a continuación:

- **datos_inscripcion**

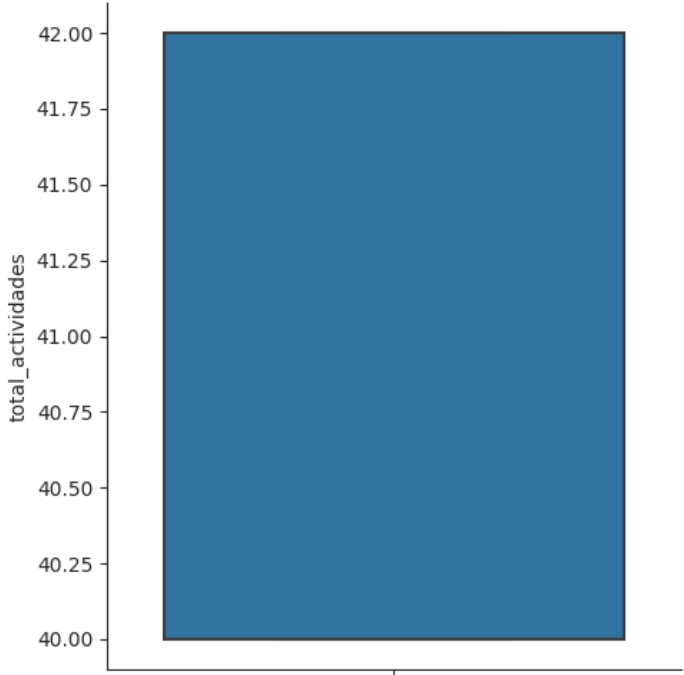
Columna / Atributo / Item	Descripción / Observaciones
id_estudiante	Se verifica el cumplimiento del formato establecido y comunicado por los expertos del negocio. Resultados: Conteo de errores en el formato del atributo id_estudiante: - Dataset: datos_inscripciones.csv => Cantidad: 0 0.0% - Dataset: datos_cursado.csv => Cantidad: 0 0.0% - Dataset: datos_academicos.csv => Cantidad: 0 0.0%
fecha_ingreso fecha_inscripcion	Se verifica el cumplimiento de la regla del negocio que establece que la fecha_inscripcion no puede ser mayor a la fecha_ingreso . Resultados: Casos de problemas detectados: 46 filas Porcentaje de filas con problemas de valores fuera de reglas de negocio: 3.45 %

- **datos_cursado**

Columna / Atributo / Item	Descripción / Observaciones
estado_inscripcion cursadas_aprobadas cursadas_regulares cursadas_libres inscripciones_examenes	Se verifica el cumplimiento de la regla del negocio que establece que un estudiante con inscripción en estado "Pendiente" no puede registrar actividad académica (<i>esto es: cursadas o exámenes</i>)

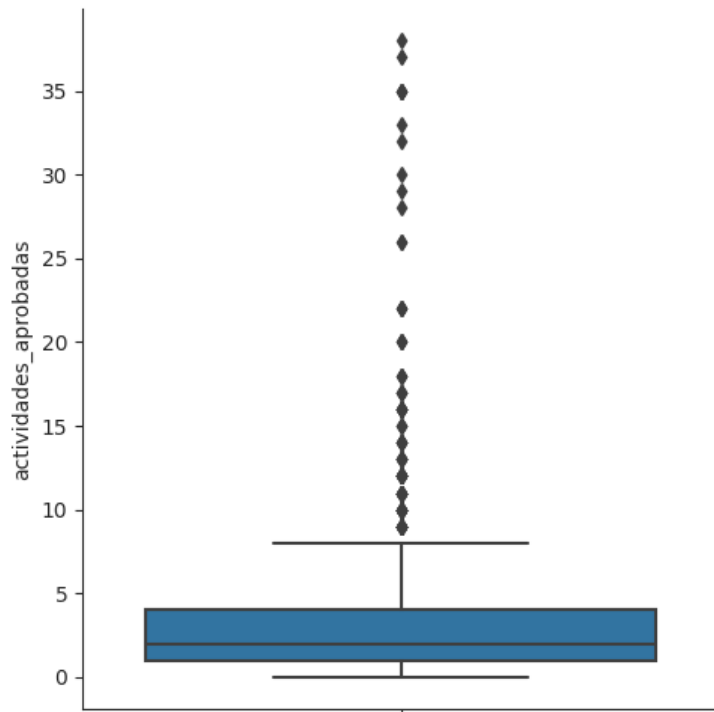
exámenes_aprobados	<p>Resultados:</p> <p>Casos de problemas de valores por reglas: 12 filas Porcentaje de filas con problemas de valores fuera de reglas de negocio: 0.9 %</p>
--------------------	--

▪ **datos_academicos**

Columna / Atributo / Item	Descripción / Observaciones
calidad regular	<p>Se verifica la regla del negocio que establece que no pueden darse las siguientes combinaciones de valores para estos atributos:</p> <ul style="list-style-type: none"> • <i>regular = N calidad = A</i> • <i>regular = S calidad = P</i> <p>Resultados:</p> <p>Casos de problemas de valores por reglas: 197 filas Porcentaje de filas con problemas de valores fuera de reglas de negocio: 14.79 %</p>
total_actividades	<p>Se verifica si existen outliers en los datos de la cantidad de asignaturas, a través de una visualización:</p>  <p>Resultados:</p> <p>Se establece que este dato no tiene valores fuera de rango</p>

actividades_aprobadas

Se verifica si existen outliers en los datos de la cantidad de asignaturas aprobadas, a través de una visualización:



Resultados:

Se observa que las asignaturas aprobadas superiores a 10 se consideran como outliers y al ser datos de estudiantes de primer año se debe revisar con el experto del dominio puesto que si es poco probable que una persona vea mas de 10 materias en un solo año.

[C] Fase de preparación de los datos

- **Selección de datos**

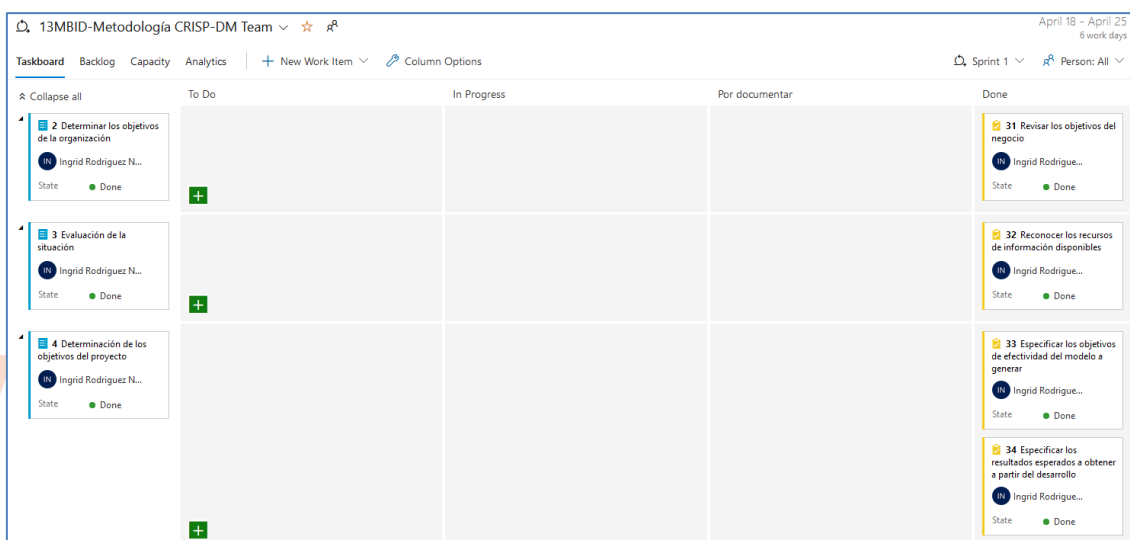
Se describe a continuación la selección de columnas realizada inicialmente en los datasets:

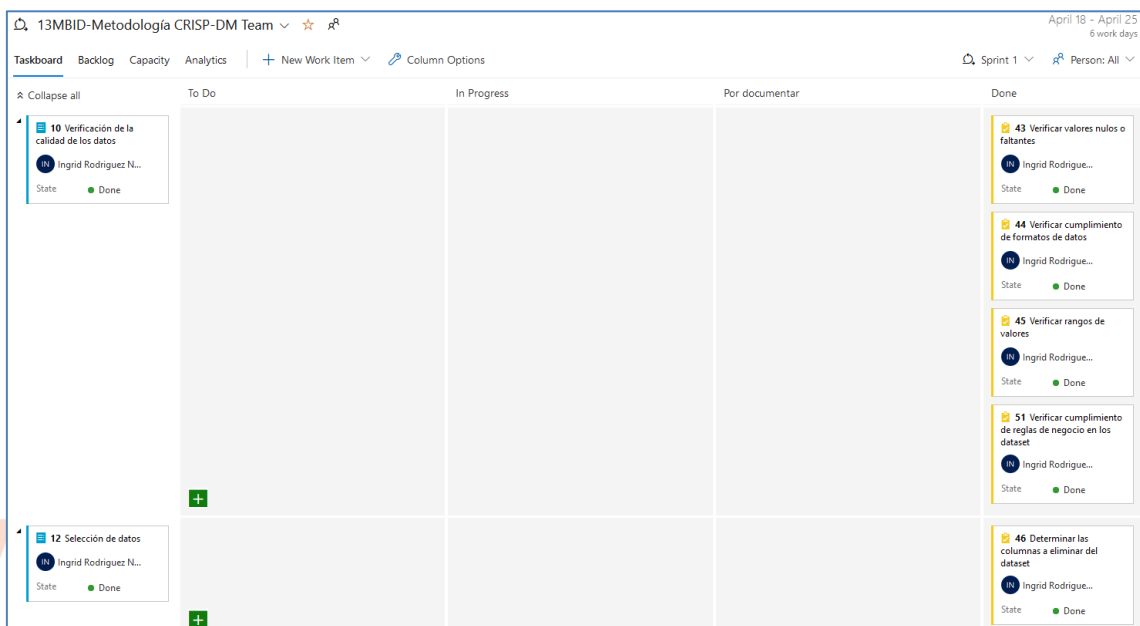
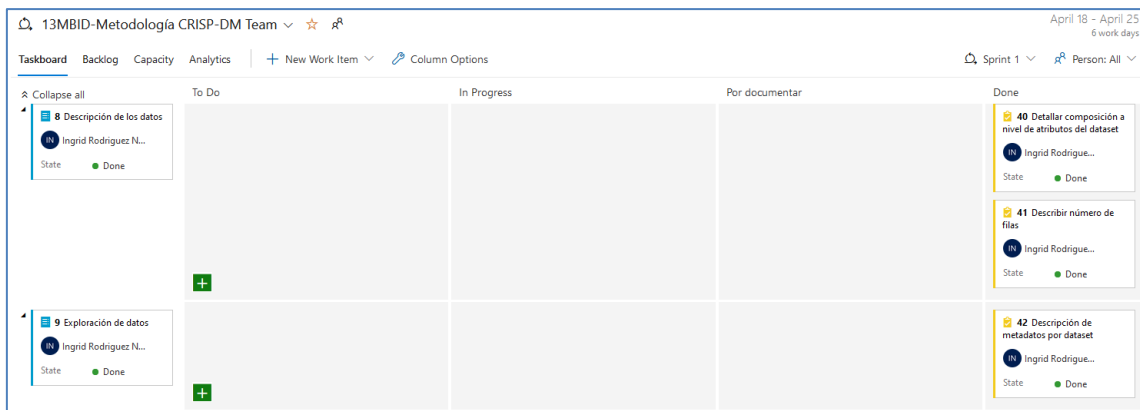
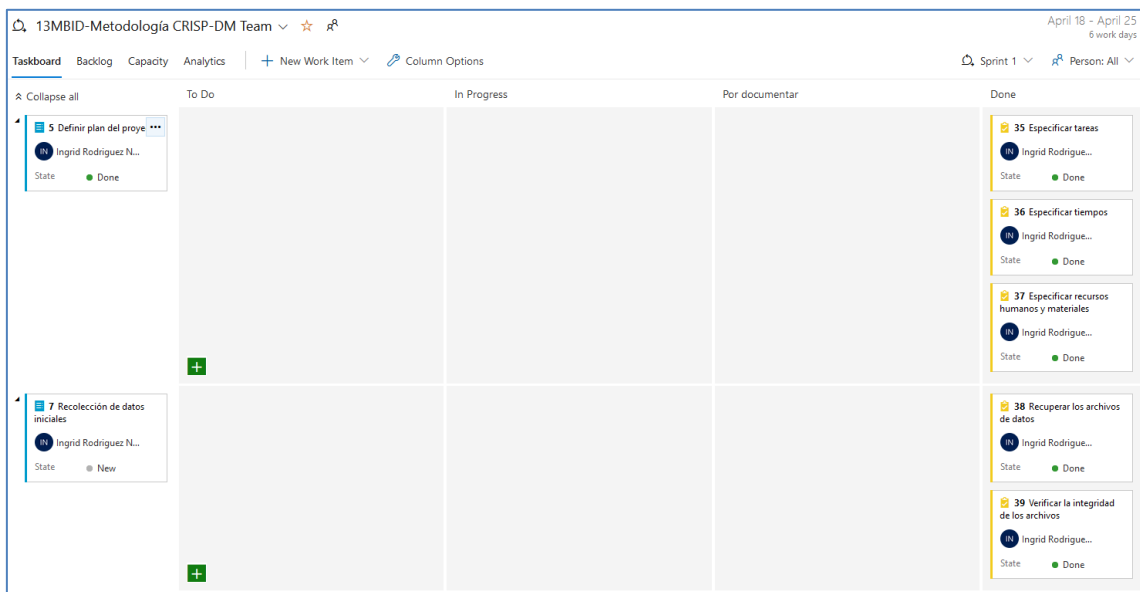
- Dataset **datos_inscripciones**, columnas eliminadas:
 - 'plan_estudios'
 - 'version_plan'
 - 'modalidad'
 - 'err_formato_matricula' (*)
 - 'regla_fechas_ingreso' (*)
- Dataset **datos_academicos**, columnas eliminadas:
 - 'plan'
 - 'fecha_ingreso'
 - 'err_formato_matricula'
 - 'regla_verificacion_calidad' (*)
- Dataset **datos_cursado**, columnas eliminadas
 - 'err_formato_matricula' (*)
 - 'regla_estado_inscripcion' (*)

(*) Se trata de columnas agregadas como parte del proceso de verificación de calidad de los datos. No se corresponden a datos del dominio.

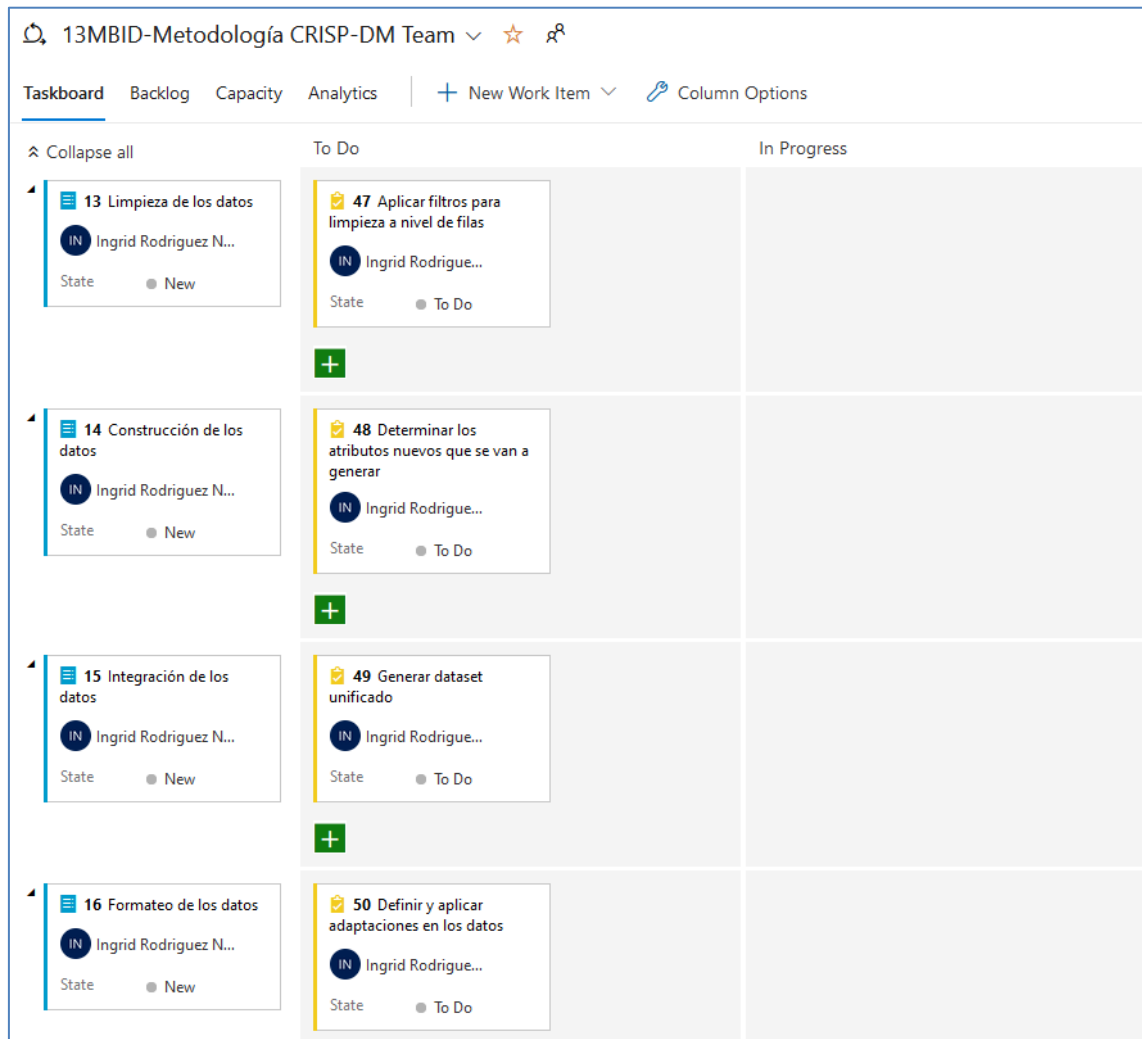
El estado del proyecto actualmente se encuentra de la siguiente manera:

- **Azure Boards – Tareas Terminadas del Sprint 1**





▪ **Azure Boards – Tareas pendientes del Sprint 1**



A partir de acá se retoma la fase de Preparación de los Datos en la actividad guiada 3.

● **Verificación de valores nulos**

En los datasets datos_inscripcion y datos_cursado no se han encontrado filas nulas.

En el dataset datos_academicos los siguientes atributos presentan valores nulos:

- fecha_ultimo_examen - **406** filas nulas
- anio_ultima_reinscripcion - **206** filas nulas

En este caso se ha definido con los expertos en el dominio no proceder con la eliminación de estas filas dado que son indicadores de no haber rendido un examen final y / o no haber manifestado la voluntad de continuar cursando por parte del estudiante.

- **Limpieza de los datos**

Se han realizado las siguientes operaciones de eliminación de filas:

- **[a] datos_inscripcion**
- **[b] datos_cursado**
- **[c] datos_academicos**

Dataset	Atributo / Columna	Observaciones / Resultados
[a]	Se eliminan las filas que no cumplen con las condiciones de la regla de verificación de fechas de ingreso / inscripción. Atributo utilizado: regla_fechas_ingreso Valor de filtro: 'err'	El dataset queda conformado por: 1286 filas. Total original: 1332 Diferencia: -46 (3.5%)
[b]	Se eliminan las filas que no cumplen con las condiciones de la regla de verificación del estado de inscripción de la persona y su condición como estudiante. Atributo utilizado: regla_estado_inscripcion Valor del filtro: 'err'	El dataset queda conformado por: 1320 filas Total original: 1332 Diferencia: -12 (0.9%)
[c]	Se eliminan las filas que no cumplen con las condiciones de la regla de verificación de la calidad de los estudiantes con respecto al resto de los ítems de análisis. Atributo utilizado: regla_verificacion_calidad Valor del filtro: 'err'	El dataset queda conformado por: 618 filas Total original: 815 Diferencia: -197 (24.3%)

- **Construcción de datos**

Se ha generado el siguiente atributo:

- **“pct_avance_carrera”** [datos_academicos]: se calcula con la siguiente fórmula:
 - $\text{actividades_aprobadas} / \text{total_actividades}$
- **“pct_asignaturas_perdidas”** [datos_cursados]: se calcula con la siguiente fórmula:
 - $\text{cursadas_libres} / \text{cursadas_totales}$

- **Integración de los datos**

Se comienza por la integración de los tres datasets originales, con los cambios que se hayan registrado hasta el momento.

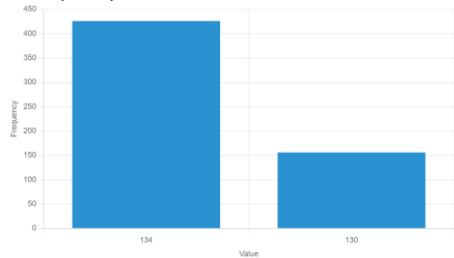
Resultados obtenidos:

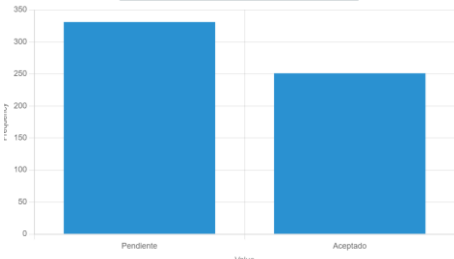
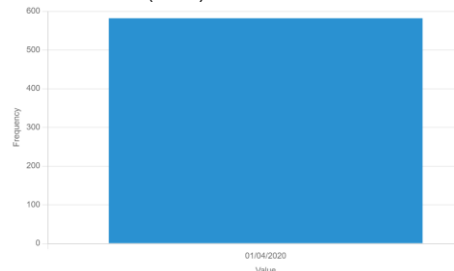
- Datos de inscripciones: **1286** - Datos de cursado: **1320** - Coincidencias entre inscripciones y cursado: **1275**
- Datos completos: **582** integrando **todas las fuentes**

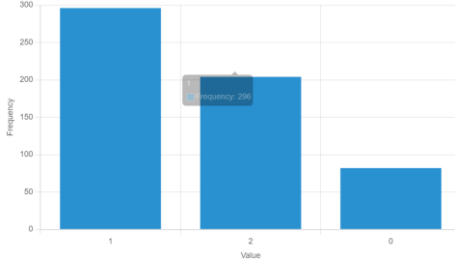
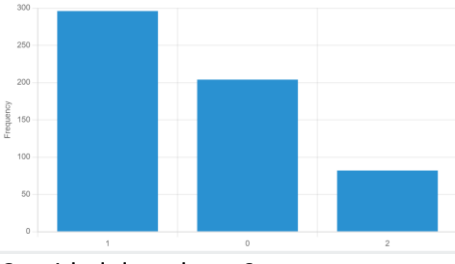

Se observa que se han perdido en la unión una gran cantidad de filas al realizar la integración con el dataset de datos_academicos. Se recomienda elevar estos resultados y consultar sobre la situación a los expertos en la gestión de los sistemas base.

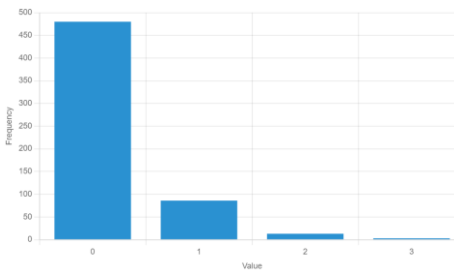
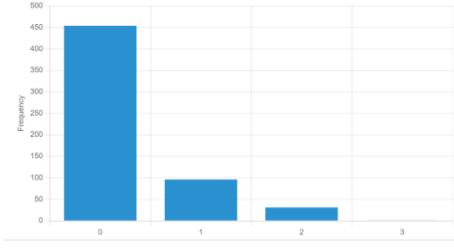
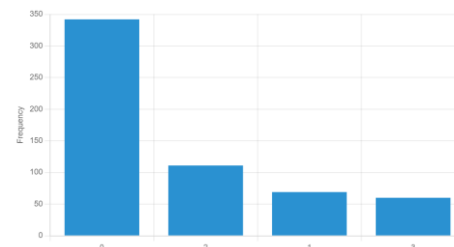
El dataset final tiene las siguientes dimensiones: **24 columnas y 582 filas**

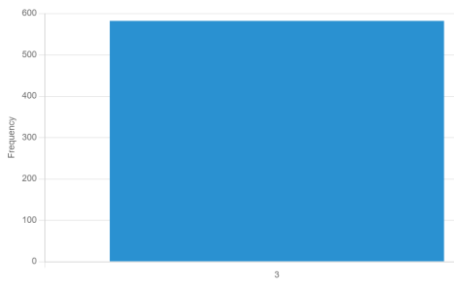
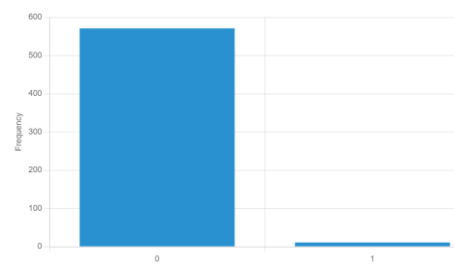
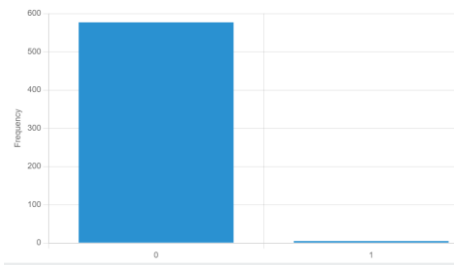
Registro de **meta-datos** del dataset completo:

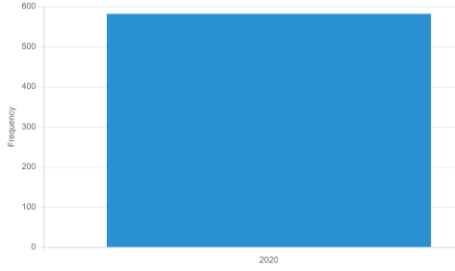
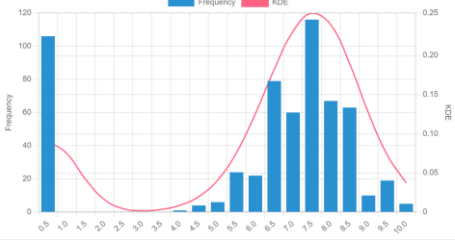
Dataset	Columna / Atributo	Tipo de datos	Observaciones
[completo]	id_estudiante	String / Alfanumérico	<p>Atributo con un formato especial: ##-#####-#</p> <p>Valores de ejemplo: CA-003269-2, CA-005300-2 (1), CA-006491-8 (1), CA-007459-0 (1), CA-008313-1 (1)</p> <p>Hace referencia al número de matrícula de cada estudiante.</p> <p>Cantidad de nulos = 0</p>
[completo]	propuesta	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>134 (426) 130 (156)</p>  <p>Cantidad de nulos = 0</p>
[completo]	estado_inscripcion	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p>

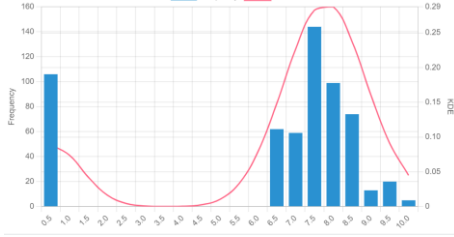
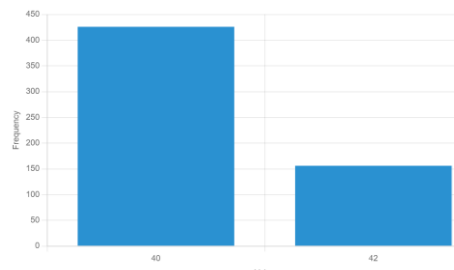
			<p>Pendiente (331) Aceptado (251)</p>  <p>Cantidad de nulos = 0</p>
[completo]	fecha_ingreso	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>01/04/2020 (582)</p>  <p>Formato de fecha: <i>dd/mm/aaaa</i> Cantidad de nulos = 0</p>
[completo]	fecha_inscripcion	String / Alfanumérico	<p>Formato de fecha: <i>dd/mm/aaaa</i></p> <p>La cantidad de valores diferentes es: 49</p> <p>Los más frecuentes son: 04/02/2020 (43) 18/12/2019 (39) 11/12/2019 (30) 17/12/2019 (30) 03/12/2019 (27)</p> <p>Cantidad de nulos = 0</p>
[completo]	ingreso_aprobadas	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>1 (296) 2 (204) 0 (82)</p>

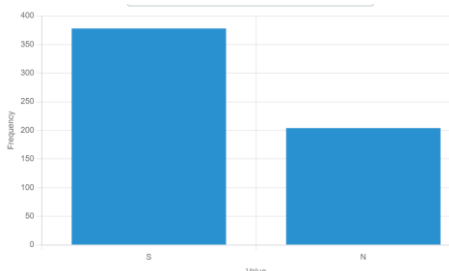
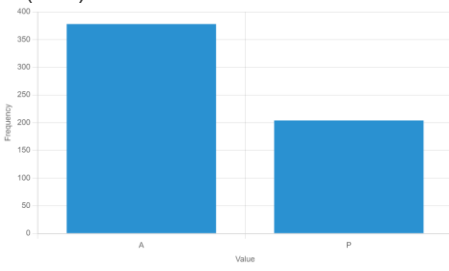
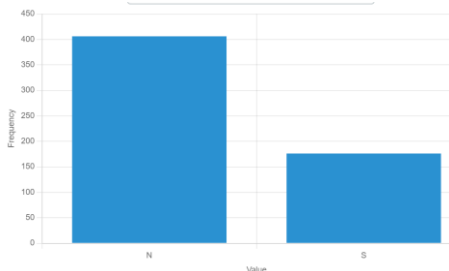
			 <p>Cantidad de nulos = 0</p>
[completo]	ingreso_libres	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>1 (296) 0 (204) 2 (82)</p>  <p>Cantidad de nulos = 0</p>
[completo]	ingreso_totales	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>2 (582)</p>  <p>Cantidad de nulos = 0</p>
[completo]	cursadas_aprobadas	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (480) 1 (86) 2 (13) 3 (3)</p>

			 <p>Cantidad de nulos = 0</p>
[completo]	cursadas_regulares	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (454) 1 (96) 2 (31) 3 (1)</p>  <p>Cantidad de nulos = 0</p>
[completo]	cursadas_libres	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (342) 2 (111) 1 (69) 3 (60)</p>  <p>Cantidad de nulos = 0</p>
[completo]	cursadas_totales	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>3 (582)</p>

			 <p>Cantidad de nulos = 0</p>
[completo]	inscripciones_examenes	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (571) 1 (11)</p>  <p>Cantidad de nulos = 0</p>
[completo]	exámenes_aprobados	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>0 (577) 1 (5)</p>  <p>Cantidad de nulos = 0</p>
[completo]	año_ingreso	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <p>2020 (582)</p>

			 <p>Cantidad de nulos = 0</p>
[completo]	fecha_ultimo_examen	String / Alfanumérico	<p>La cantidad de valores diferentes es: 22</p> <p>Los más frecuentes son: 18/08/2021 (46) 20/08/2021 (39) 23/08/2021 (34) 19/08/2021 (15) 05/08/2021 (11)</p> <p>Cantidad de nulos = 380 (65.29%) Datos con valor = 219 (34.71%)</p>
[completo]	anio_ultima_reinscripcion	float	<p>Descripción y distribución de valores (cantidad):</p> <p>2021.0 (378)</p> <p>Cantidad de nulos = 204 (35.05%) Datos con valor = 378 (64.95%)</p>
[completo]	promedio_sin_aplares	float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (10)</p>  <p>La mayoría de las notas de los estudiantes están entre 6.5 y 8.5</p> <p>Cantidad de nulos = 0</p>
[completo]	promedio_con_aplares	float	<p>Descripción y distribución de valores (cantidad):</p>

			<p>Rango de valores: (0) a (10)</p>  <p>La mayoría de las notas de los estudiantes están entre 6.5 y 8.5</p> <p>Cantidad de nulos = 0</p>
[completo]	actividades_aprobadas	Integer / Numérico	<p>La cantidad de valores diferentes es: 30</p> <p>Los más frecuentes son (cantidad):</p> <ul style="list-style-type: none"> 0 (106) 1 (98) 2 (86) 3 (72) 4 (44) 5 (32) <p>Cantidad de nulos = 0</p>
[completo]	total_actividades	Integer / Numérico	<p>Descripción y distribución de valores (cantidad):</p> <ul style="list-style-type: none"> 40 (426) 42 (156)  <p>Cantidad de nulos = 0</p>
[completo]	regular	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <ul style="list-style-type: none"> S (378) N (204)

			 <p>Cantidad de nulos = 0</p>
[completo]	calidad	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>A (378) P (204)</p>  <p>Cantidad de nulos = 0</p>
[completo]	segundo_anio	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>N (406) S (176)</p>  <p>Cantidad de nulos = 0</p>

- **Formateo de los datos**

Esta operación va a ser realizada en forma previa al inicio del proceso de generación de modelos, más exactamente en la **“Adaptación de los datos”**, por favor ir a la siguiente fase para poder visualizarlo.

Se finaliza el Sprint 1 con todas las tareas terminadas (Done):

🔍 13MBID-Metodología CRISP-DM Team ▾ ☆ 🔊

Taskboard **Backlog** Capacity Analytics | + New Work Item 🔗 Column Options ...

+	☐	Order	Title	State	Assigned To
+		1	▾ Determinar los objetivos de la organización	... ● Done	Ingrid Rodriguez Nova
			Revisar los objetivos del negocio	● Done	Ingrid Rodriguez Nova
		2	▾ Evaluación de la situación	● Done	Ingrid Rodriguez Nova
			Reconocer los recursos de información disponibles	● Done	Ingrid Rodriguez Nova
		3	▾ Determinación de los objetivos del proyecto	● Done	Ingrid Rodriguez Nova
			Especificar los objetivos de efectividad del modelo a ge...	● Done	Ingrid Rodriguez Nova
			Especificar los resultados esperados a obtener a partir ...	● Done	Ingrid Rodriguez Nova
		4	▾ Definir plan del proyecto	● Done	Ingrid Rodriguez Nova
			Especificar tareas	● Done	Ingrid Rodriguez Nova
			Especificar tiempos	● Done	Ingrid Rodriguez Nova
			Especificar recursos humanos y materiales	● Done	Ingrid Rodriguez Nova
		5	▾ Recolección de datos iniciales	● Done	Ingrid Rodriguez Nova
			Recuperar los archivos de datos	● Done	Ingrid Rodriguez Nova
			Verificar la integridad de los archivos	● Done	Ingrid Rodriguez Nova
		6	▾ Descripción de los datos	● Done	Ingrid Rodriguez Nova
			Detallar composición a nivel de atributos del dataset	● Done	Ingrid Rodriguez Nova
			Describir número de filas	● Done	Ingrid Rodriguez Nova
		7	▾ Exploración de datos	● Done	Ingrid Rodriguez Nova
			Descripción de metadatos por dataset	● Done	Ingrid Rodriguez Nova

13MBID-Metodología CRISP-DM Team				
Taskboard Backlog Capacity Analytics + New Work Item Column Options ...				
Order	Title	State	Assigned To	
8	Verificación de la calidad de los datos	Done	Ingrid Rodriguez Nova	
	Verificar valores nulos o faltantes	Done	Ingrid Rodriguez Nova	
	Verificar cumplimiento de formatos de datos	Done	Ingrid Rodriguez Nova	
	Verificar rangos de valores	Done	Ingrid Rodriguez Nova	
	Verificar cumplimiento de reglas de negocio en los data...	Done	Ingrid Rodriguez Nova	
9	Selección de datos	Done	Ingrid Rodriguez Nova	
	Determinar las columnas a eliminar del dataset	Done	Ingrid Rodriguez Nova	
10	Limpieza de los datos	Done	Ingrid Rodriguez Nova	
	Aplicar filtros para limpieza a nivel de filas	Done	Ingrid Rodriguez Nova	
11	Construcción de los datos	Done	Ingrid Rodriguez Nova	
	Determinar los atributos nuevos que se van a generar	Done	Ingrid Rodriguez Nova	
12	Integración de los datos	Done	Ingrid Rodriguez Nova	
	Generar dataset unificado	Done	Ingrid Rodriguez Nova	
	Registrar metadatos actualizados del nuevo dataset	Done	Ingrid Rodriguez Nova	
13	Formateo de los datos	Done	Ingrid Rodriguez Nova	
	Definir y aplicar adaptaciones en los datos	Done	Ingrid Rodriguez Nova	

En la siguiente vista se ven con claridad las épicas terminadas y las que quedarían para el Sprint 2:

13MBID-Metodología CRISP-DM Team

Board

Analytics

View as Backlog

New

<

In Progress

0/5

Done

<

New item

17 Modelado

IN

Ingrid Rodriguez Nova

State

New

18 Evaluación

IN

Ingrid Rodriguez Nova

State

New

19 Despliegue

IN

Ingrid Rodriguez Nova

State

New

11 Preparación de los datos

IN

Ingrid Rodriguez Nova

State

Done

6 Comprensión de los datos

IN

Ingrid Rodriguez Nova

State

Done

1 Comprensión del negocio

IN

Ingrid Rodriguez Nova

State

Done

A continuación, se muestra como quedo planeados el **Sprint 2** en Azure DevOps:

13MBID-Metodología CRISP-DM Team							
Backlog Analytics + New Work Item View as Board Column Options ...							
+ Order	Work Item T...	Title	State	Effort	Assigned To	Iteration Path	
1	Epic	Modelado	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Selección de la técnica de modelado	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Planteamiento inicial de técnicas	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Selección de técnicas aplicables	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Generar el diseño de las pruebas	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Determinar tipo y cantidad de pruebas	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Construcción del modelo	New	8	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Configuración de parámetros	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Generación de los modelos	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Descripción de los modelos	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Evaluar el modelo	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Evaluación de efectividad de cada modelo	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Determinar ítems a analizar de la config...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
2	Epic	Evaluación	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Evaluar los resultados	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Evaluar resultados con los criterios del c...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Aprobar o descartar modelos	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Revisar el proceso	New	5	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Revisión y definición de mejoras aplicabl...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Determinar pasos siguientes	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Listar acciones de mejora aplicables	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Determinar acciones en la próxima itera...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	

13MBID-Metodología CRISP-DM Team							
Backlog Analytics + New Work Item View as Board Column Options ...							
+ Order	Work Item T...	Title	State	Effort	Assigned To	Iteration Path	
3	Epic	Despliegue	New		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Planificar el despliegue	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Determinar esquema de implementación	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Planificar monitoreo y mantenimiento	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Plan de actualización y mantenimiento	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Determinar e implementar acciones del ...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Producir el reporte final	New	2	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Construcción del reporte final	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Presentación del reporte	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Product Ba...	Revisar el proyecto	New	3	Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	
	Task	Documentación de conocimiento adquiri...	To Do		Ingrid Rodriguez Nova	13MBID-Metodología CRISP-DM\Sprint 2	

[D] Fase de modelado

- **Selección de la técnica de modelado**

Con base en los objetivos planteados al inicio del proyecto, las técnicas a utilizar para generar el/los modelos que conformarán el producto de datos final son:

- Árboles de decisión
- Redes neuronales
- Regresión Logística
- Vecinos mas Cercanos
- Métodos de ensamblado de modelos

Además, se van a considerar técnicas o métodos de gestión de parámetros para las técnicas involucradas, así como también para la fase de evaluación de los resultados de cada modelo.

- **Adaptación de los datos**

Una vez iniciado el trabajo de la fase de Modelado se ha determinado realizar las siguientes modificaciones sobre el dataset disponible:

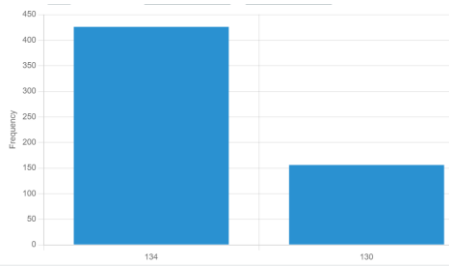
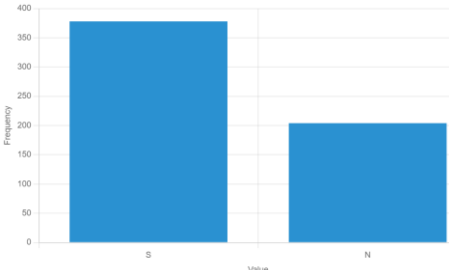
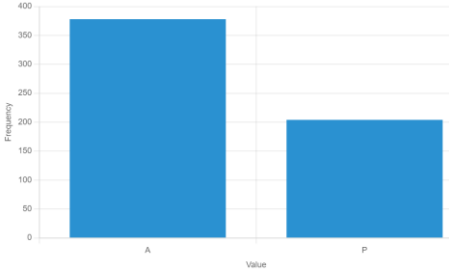
Operación	Atributo/s	Descripción / Observaciones
Generación de un atributo	avance_ingreso (función: pct_avance_ingreso)	Marca el % de avance logrado en las asignaturas de ingreso a la carrera. Fórmula: - ingreso_aprobadas / ingreso_totales
Generación de un atributo	avance_semestre (función: pct_avance_semestre)	Marca el % de asignaturas aprobadas en el cursado del 1er semestre de la carrera. Fórmula: - (cursadas_aprobadas + cursadas_regulares) / cursadas_totales
Generación de un atributo	avance_carrera (función: pct_avance_carrera)	Marca el % del avance del estudiante en la carrera. Fórmula: actividades_aprobadas / total_actividades

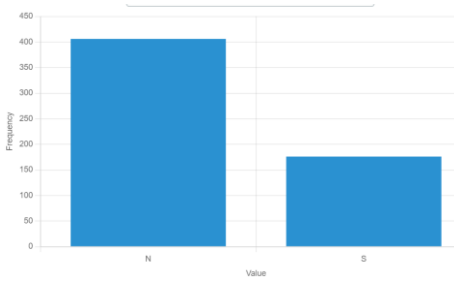
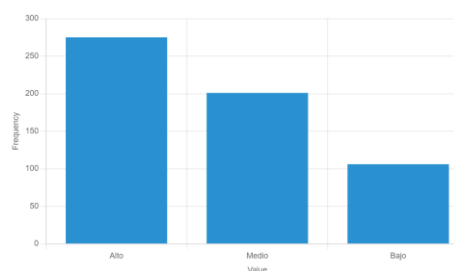
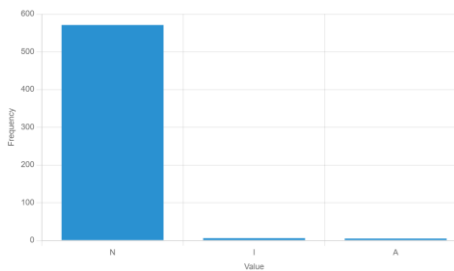
Generación de un atributo	asignaturas_perdidas (función: pct_ asignaturas_perdidas)	Marca el % de asignaturas pérdidas durante un semestre. Fórmula: cursadas_libres/ cursadas_totales
Generación de un atributo	exámenes_1er_semestre (función: exámenes_1er_semestre)	Indica si un estudiante se inscribió, aprobó algún examen o nunca asistió. Fórmula: Si inscripciones_exámenes > 0 & exámenes_aprobados > 0 Entonces A (Aprobado) Si No y Si inscripciones_exámenes > 0 Entonces I (Inscrito) Si No N (Nunca asistió a la carrera).
Generación de un atributo	rango_promedios	Se crea un atributo para establecer por categorías de Alto, Medio y Bajo el promedio de las calificaciones con aplazamientos. Formula: De 0 a 5 = Bajo De 5 a 7 = Medio De 7 a 10 = Alto
Generación de un atributo	rindio_examen	Obtiene sus valores a partir del atributo 'fecha_ultimo_examen': -Si es una fecha nula = "N" -Si es una fecha concreta = "S" indicando que sí ha rendido exámenes
Generación de un atributo	inscripto_ult_ciclo	Obtiene sus valores a partir del atributo 'anio_ultima_reinscripcion': -Si el valor es 2021 = "S" -En caso de valor diferente o nulo = "N" indicando que no se

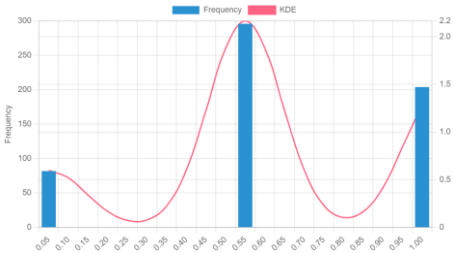
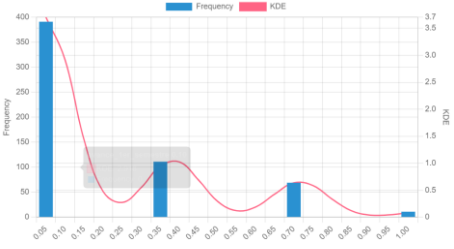
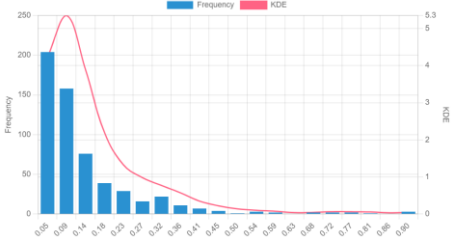
		ha inscrito a cursar nuevamente
Eliminación de atributos	'actividades_aprobadas' 'total_actividades' 'ingreso_aprobadas' 'ingreso_libres' 'ingreso_totales' 'cursadas_aprobadas' 'cursadas_regulares' 'cursadas_libres' 'cursadas_totales' 'fecha_ingreso' 'fecha_inscripcion' 'anio_ingreso' 'inscripciones_exámenes' 'exámenes_aprobados' 'promedio_sin_aplazos' 'promedio_con_aplazos' 'id_estudiante' 'estado_inscripcion' 'fecha_ultimo_examen' 'anio_ultima_reinscripcion'	Se han eliminado los atributos utilizados para la generación de las columnas nuevas mencionadas en esta misma tabla. Además de otros atributos que presentan valores constantes y/o no son requeridos para el tipo de análisis que se desea realizar.
Actualización del atributo	estado_inscripcion por 'estadoN'	Se han transformado sus valores según el siguiente criterio: ○ "Pendiente" = "P" ○ "Aceptada" = "A"

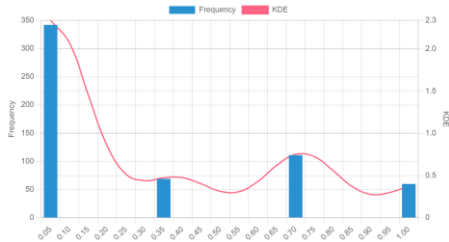
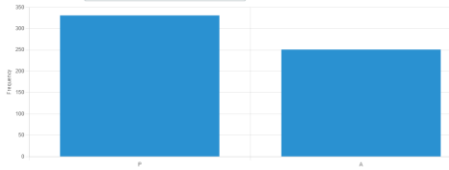
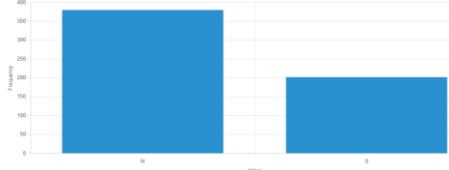
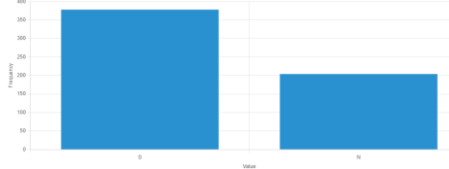
A continuación, se establecen los **meta-datos** del dataset [filtrado] como resultado de la adaptación de los datos:

Dataset	Columna / Atributo	Tipo de datos	Observaciones
[filtrado]	propuesta	Integer / Numérico	Descripción y distribución de valores (cantidad): 134 (426) 130 (156)

			 <p>Cantidad de nulos = 0</p>
[filtrado]	regular	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>S (378) N (204)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	calidad	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>A (378) P (204)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	segundo_anio	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>N (406) S (176)</p>

			 <p>Cantidad de nulos = 0</p>
[filtrado]	rango_promedios	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>Alto (275) Medio (201) Bajo (106)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	exámenes_1er_sem estre	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>N (571) I (6) A (5)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	avance_ingreso	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (1)</p>

			 <p>La mayoría de los estudiantes tuvieron un 55% de avance en las asignaturas ingresadas</p> <p>Cantidad de nulos = 0</p>
[filtrado]	avance_1er_semestre	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (1)</p>  <p>La cantidad de los estudiantes que tienen un 100% de avance en el semestre son muy pocos. Cerca de 400 estudiantes no avanzaron nada en el semestre</p> <p>Cantidad de nulos = 0</p>
[filtrado]	avance_carrera	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (1)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	asignaturas_perdidas	Float	<p>Descripción y distribución de valores (cantidad):</p> <p>Rango de valores: (0) a (1)</p>

			 <p>Cerca de 100 estudiantes tienen el 70% de las asignaturas perdidas.</p> <p>Cantidad de nulos = 0</p>
[filtrado]	estadoN	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>P (331) - [Pendiente] A (251) - [Aceptada]</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	rindio_examen	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>N (380) S (202)</p>  <p>Cantidad de nulos = 0</p>
[filtrado]	inscripto_ult_ciclo	String / Alfanumérico	<p>Descripción y distribución de valores (cantidad):</p> <p>S (378) N (204)</p>  <p>Cantidad de nulos = 0</p>

- **Generación del plan de pruebas**

En primer lugar, a nivel de distribución de las filas del dataset procesado hasta este punto se ha optado por trabajar de la siguiente manera:

- Se utilizará un 75% de los datos para tareas de entrenamiento.
- Se utilizará el restante 25% para tareas de testeo o evaluación de modelos.

En segunda instancia, se pasará a realizar una experimentación según los siguientes criterios:

- Para cada técnica a emplear se documentarán sus parámetros de entrenamiento con el dataset disponible y, una vez obtenidos los resultados, se registrará la efectividad de su clasificación sobre el dataset de testeo.
- Esto será repetido un mínimo de tres (3) iteraciones a través de las cuales se irán seleccionando aquellas técnicas con un mejor rendimiento. Al pasar de una etapa a otra se podrán realizar modificaciones en los parámetros para optimizar los resultados obtenidos.

- **Construcción del Modelo**

En esta oportunidad se van a utilizar librerías implementadas sobre el lenguaje python que proveen diferentes métodos de machine learning para realizar el procesamiento de los datos y la generación de los modelos. El código de estas actividades se encuentra compilado en libretas de jupyter disponibles en el siguiente repositorio: <https://github.com/JulietaNova/13MBID-AG3-Repositorio>

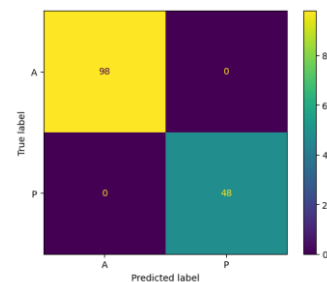
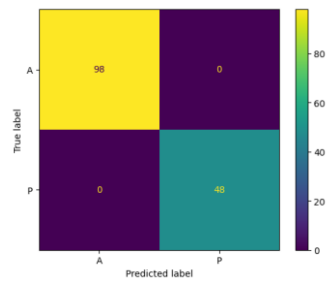
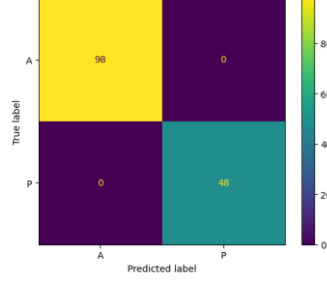
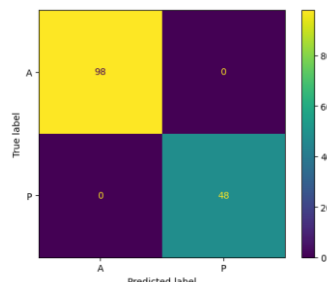
Los resultados de la experimentación serán resumidos en las próximas secciones de este documento.

- **Evaluación del modelo**

Prueba #1

Técnica	Parámetros ¹	Resultados
LogisticRegression	LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=100, multi_class='auto', n_jobs=None, penalty='l2', random_state=None, solver='liblinear' , tol=0.0001, verbose=0, warm_start=False)	Rendimiento obtenido: 1.0 Matriz de confusión:

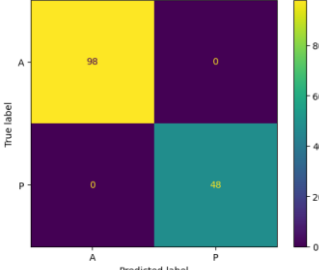
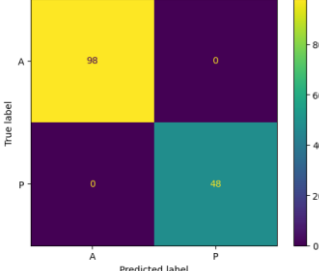
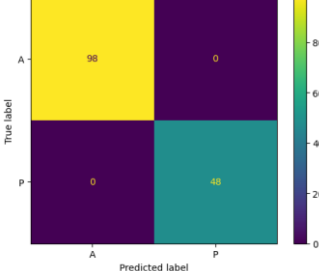
¹ Los nombres y valores de parámetros que se encuentran **destacados** han sido modificados con respecto a su valor por defecto en la ejecución de la técnica.

		
KNeighborsClassifier	KNeighborsClassifier(algorithm='ball_tree' , leaf_size=25 , metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=50 , p=2, weights='uniform')	<p>Rendimiento obtenido: 1.0</p> <p>Matriz de confusión:</p> 
DecisionTreeClassifier	DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy' , max_depth=3 , max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1 , min_samples_split=10, min_weight_fraction_leaf=0.0, random_state=None, splitter='best')	<p>Rendimiento obtenido: 1.0</p> <p>Matriz de confusión:</p> 
RandomForestClassifier	RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=None, max_features='sqrt', max_leaf_nodes=None, max_samples=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=None, oob_score=False, random_state=None, verbose=0, warm_start=False)	<p>Rendimiento obtenido: 1.0</p> <p>Matriz de confusión:</p> 

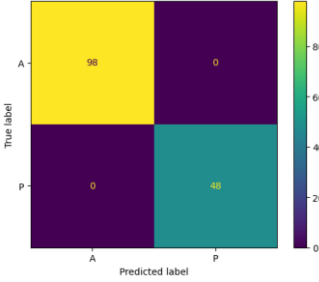
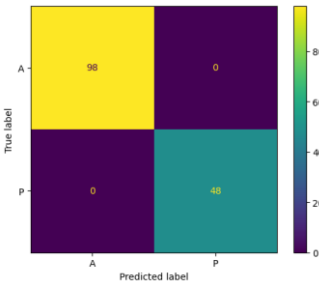
GradientBoostingClassifier	<pre>GradientBoostingClassifier(ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=1.0, loss='log_loss', max_depth=1, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=100, n_iter_no_change=None, random_state=0, subsample=1.0, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False)</pre>	<p>Rendimiento obtenido: 1.0</p> <p>Matriz de confusión:</p> <table border="1"> <thead> <tr> <th></th> <th>Predicted label A</th> <th>Predicted label P</th> </tr> </thead> <tbody> <tr> <th>True label A</th> <td>98</td> <td>0</td> </tr> <tr> <th>True label P</th> <td>0</td> <td>48</td> </tr> </tbody> </table>		Predicted label A	Predicted label P	True label A	98	0	True label P	0	48
	Predicted label A	Predicted label P									
True label A	98	0									
True label P	0	48									

Prueba #2

Se elimina el atributo [regular] y se mantienen los parámetros de la Prueba 1 para realizar las pruebas #2.

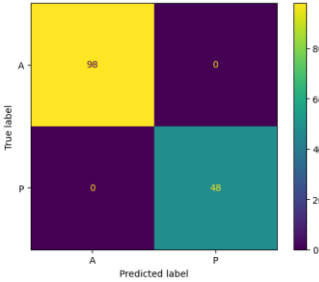
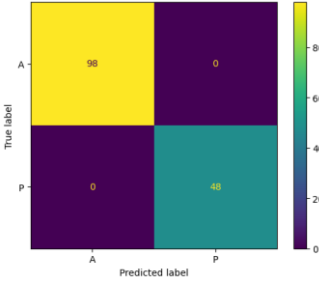
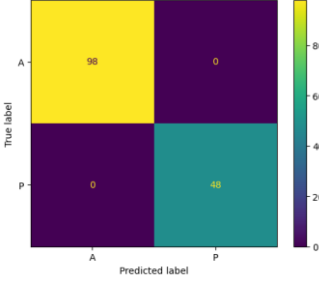
Técnica	Parámetros ²	Resultados
LogisticRegression	LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=100, multi_class='auto', n_jobs=None, penalty='l2', random_state=None, solver='liblinear' , tol=0.0001, verbose=0, warm_start=False)	Rendimiento obtenido: 1.0 Matriz de confusión: 
KNeighborsClassifier	KNeighborsClassifier(algorithm='ball_tree' , leaf_size=25 , metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=50 , p=2, weights='uniform')	Rendimiento obtenido: 1.0 Matriz de confusión: 
DecisionTreeClassifier	DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy' , max_depth=3 , max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1 , min_samples_split=10, min_weight_fraction_leaf=0.0, random_state=None, splitter='best')	Rendimiento obtenido: 1.0 Matriz de confusión: 
RandomForestClassifier	RandomForestClassifier(bootstrap=True, ccp_alpha=0.0,	Rendimiento obtenido: 1.0

² Los nombres y valores de parámetros que se encuentran **destacados** han sido modificados con respecto a su valor por defecto en la ejecución de la técnica.

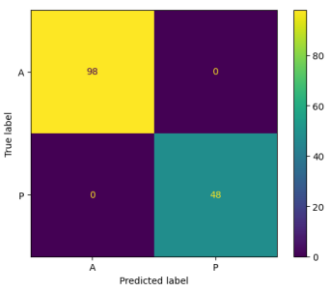
	<pre>class_weight=None, criterion='gini', max_depth=None, max_features='sqrt', max_leaf_nodes=None, max_samples=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=None, oob_score=False, random_state=None, verbose=0, warm_start=False)</pre>	<p>Matriz de confusión:</p> 
GradientBoostingClassifier	<pre>GradientBoostingClassifier(ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=1.0, loss='log_loss', max_depth=1, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=100, n_iter_no_change=None, random_state=0, subsample=1.0, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False)</pre>	<p>Rendimiento obtenido: 1.0</p> <p>Matriz de confusión:</p> 

Prueba #3

Se mantiene el mismo conjunto de datos, pero modifican los parámetros en los modelos

Técnica	Parámetros ³	Resultados
LogisticRegression	LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=100, multi_class='auto', n_jobs=None, penalty='l2', random_state=None, solver='lbfgs' , tol=0.0001, verbose=0, warm_start=False)	Rendimiento obtenido: 1.0 Matriz de confusión: 
KNeighborsClassifier	KNeighborsClassifier(algorithm='ball_tree' , leaf_size=15, metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=5 , p=2, weights='distance')	Rendimiento obtenido: 1.0 Matriz de confusión: 
DecisionTreeClassifier	DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='log_loss' , max_depth=7 , max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=10 , min_weight_fraction_leaf=0.0, random_state=None, splitter='random')	Rendimiento obtenido: 1.0 Matriz de confusión: 
RandomForestClassifier	RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=None,	Rendimiento obtenido: 1.0 Matriz de confusión:

³ Los nombres y valores de parámetros que se encuentran **destacados** han sido modificados con respecto a su valor por defecto en la ejecución de la técnica.

	<pre> max_features='sqrt', max_leaf_nodes=None, max_samples=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=35, min_weight_fraction_leaf=0.0, n_estimators=40, n_jobs=None, oob_score=False, random_state=9, verbose=0, warm_start=False) </pre>	 <table border="1"> <thead> <tr> <th></th> <th>Predicted label: A</th> <th>Predicted label: P</th> </tr> </thead> <tbody> <tr> <th>True label: A</th> <td>98</td> <td>0</td> </tr> <tr> <th>True label: P</th> <td>0</td> <td>48</td> </tr> </tbody> </table>		Predicted label: A	Predicted label: P	True label: A	98	0	True label: P	0	48
	Predicted label: A	Predicted label: P									
True label: A	98	0									
True label: P	0	48									

[E] Fase de evaluación

- **Evaluación de los resultados**

En función de los resultados obtenidos en la ejecución del plan de pruebas documentado en la sección anterior se ha seleccionado la técnica **KNeighborsClassifier** y la de **LogisticRegression** para ser utilizada sobre los datos de estudiantes nuevos. Esto se ha debido a que la efectividad obtenida en las dos técnicas en la fase de entrenamiento y testeo ha sido del 100%.

- **Proceso de revisión**

En función de los resultados obtenidos, las técnicas Regresión Logística y de Vecinos más Cercanos mencionada en el paso anterior será utilizada para procesar los datos de los **estudiantes de 2022** y determinar el valor de su atributo “calidad” en base a sus datos almacenados en el sistema de gestión académica. Los resultados de esta predicción se encuentran en el apartado “**Informe final**” del presente documento.

- **Determinación de futuras tareas**

Como tareas que podrían ser de utilidad para mejorar el rendimiento del modelo de predicción generado se pueden mencionar:

- Incorporar datos referidos a la situación socioeconómica de cada estudiante, por ejemplo: si trabaja, si tiene personas a cargo, entre otros.
- Incorporar datos de los estudiantes como fecha de nacimiento para determinar la edad, año de graduación del colegio.
- Incorporar datos de si tiene otras carreras de pregrado.
- Se espera tiene un mayor número de datos porque los modelos terminaron todos con una efectividad del 100% y sabemos que esto no es comportamiento normal de los modelos.

En la **próxima iteración** del proyecto se propone ejecutar las siguientes tareas:

- Verificar cuestiones de integridad referencial en los datos disponibles para incrementar la cantidad de filas utilizadas para el entrenamiento de los modelos de aprendizaje automático.
- Aumentar el equipo en capacidades para mejorar la evaluación de los modelos, puesto que todos con una efectividad tan alta indica que o es problema de los datos o es un problema de falta de experiencia en el equipo para implementar este tipo de actividades asociadas al Científico de Datos.

[F] Fase de despliegue

- **Plan de implementación**

Las autoridades de la Universidad han determinado que el modelo generado sea utilizado como herramienta de soporte para la definición de políticas de apoyo a los estudiantes que inician sus estudios año a año.

Por tal motivo, se ha dispuesto que el modelo sea actualizado con datos nuevos cada año (*en una fecha a determinar en función de la finalización de la actividad académica de los estudiantes*) y vuelva a ser aplicado a los nuevos estudiantes una vez finalizado su primer semestre de cursado.

- **Supervisión y Mantenimiento**

Una vez que el producto se encuentre en uso por parte de sus usuarios objetivo se podrían realizar las siguientes acciones:

- Monitoreo de las predicciones que se realizan para que sean consistentes con los datos de entrenamiento y testeo a fin de detectar desvíos.
- Contabilizar accesos a la herramienta para obtener métricas de uso.
- Establecer puntos de control para verificar la degradación del modelo.
- Hacer seguimiento al tiempo de ejecución del modelo para garantizar que tiene los recursos suficientes para correr a medida del crecimiento de los datos.

- **Informe Final**

Se presentan los resultados de aplicación del modelo generado con la técnica de mejor rendimiento sobre los datos de estudiantes nuevos para el periodo 2022:

Se tuvo un error con la ejecución del notebook.

	Cantidad	Porcentaje
Alumnos Activos		
Alumnos Pasivos		
Total		

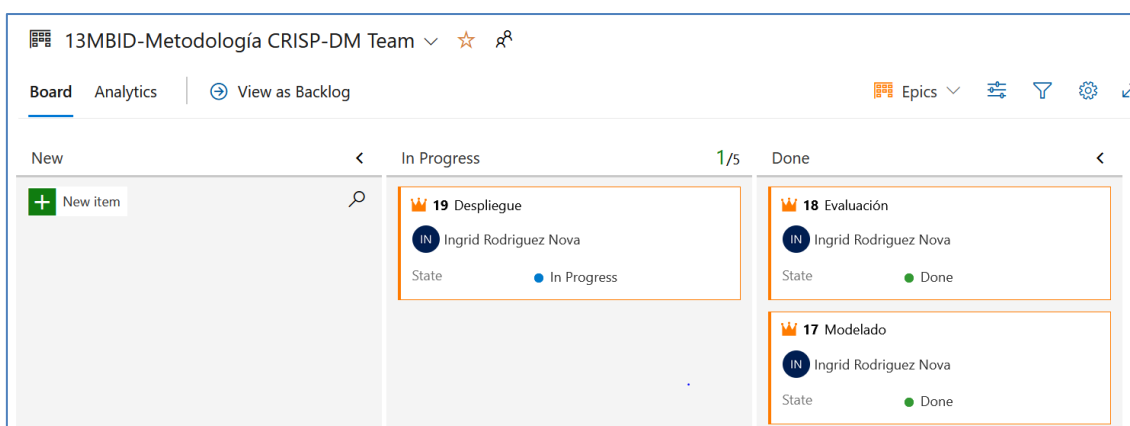
- **Revisión del proyecto**

Una vez finalizada la presente iteración de la metodología CRISP-DM se reconocen como mejoras aplicables al proyecto las siguientes:

- Incorporar herramientas que brinden soporte a la comunicación con los expertos en el dominio a fin de solucionar inquietudes del equipo de trabajo con mayor velocidad para que el equipo pueda entender el negocio.
- Fortalecer los conocimientos para que el equipo desarrolle de mejor manera y con mayor velocidad.

A continuación, se presenta como se finalizó el Sprint 2:

Se tuvo un problema para terminar la tarea de “Producir el reporte final”.



Las tareas en específico se establecen las tareas en seguida:

13MBID-Metodología CRISP-DM Team

July 1 - July 14

1 work day remaining

Taskboard

Backlog

Capacity

Analytics

+ New Work Item

Column Options

...

Sprint 2

<div><div></div><div></div></div>	Order	Title		State	Assigned To	Remai...
<div><div></div><div></div></div>	1	<div><div></div>Selección de la técnica de modelado</div>	<div><div></div><div></div></div>	<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Planteamiento inicial de técnicas</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Selección de técnicas aplicables</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Adaptación de los datos</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
2		<div><div></div>Generar el diseño de las pruebas (Plan de pruebas)</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Determinar tipo y cantidad de pruebas</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
3		<div><div></div>Construcción del modelo</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Configuración de parámetros</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Generación de los modelos</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	
		<div><div></div>Descripción de los modelos</div>		<div><div></div>Done</div>	Ingrid Rodriguez Nova	

🔍 13MBID-Metodología CRISP-DM Team ▾ ☆ 👤 July 1 - July 14
1 work day remaining

Taskboard **Backlog** Capacity Analytics | + New Work Item 🔗 Column Options ... 🔄 Sprint 2 ▾

Order	Title	State	Assigned To	Remai...
4	<div> <div>▼</div> <div> <div>📋</div> <div>Evaluar el modelo</div> </div> </div> <div> <div>📌</div> <div>Evaluación de efectividad de cada modelo</div> </div> <div> <div>📌</div> <div>Determinar ítems a analizar de la configuración</div> </div>	<div>● Done</div> <div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
5	<div> <div>▼</div> <div> <div>📋</div> <div>Evaluar los resultados</div> </div> </div> <div> <div>📌</div> <div>Evaluar resultados con los criterios del cliente</div> </div> <div> <div>📌</div> <div>Aprobar o descartar modelos</div> </div>	<div>● Done</div> <div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
6	<div> <div>▼</div> <div> <div>📋</div> <div>Revisar el proceso</div> </div> </div> <div> <div>📌</div> <div>Revisión y definición de mejoras aplicables al proceso</div> </div>	<div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
7	<div> <div>▼</div> <div> <div>📋</div> <div>Determinar pasos siguientes</div> </div> </div> <div> <div>📌</div> <div>Listar acciones de mejora aplicables</div> </div> <div> <div>📌</div> <div>Determinar acciones en la próxima iteración</div> </div>	<div>● Done</div> <div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	

🔍 13MBID-Metodología CRISP-DM Team ▾ ☆ 👤 July 1 - July 14
1 work day remaining

Taskboard **Backlog** Capacity Analytics | + New Work Item 🔗 Column Options ... 🔄 Sprint 2 ▾

Order	Title	State	Assigned To	Remai...
	<div> <div>📌</div> <div>Listar acciones de mejora aplicables</div> </div> <div> <div>📌</div> <div>Determinar acciones en la próxima iteración</div> </div>	<div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
8	<div> <div>▼</div> <div> <div>📋</div> <div>Planificar el despliegue</div> </div> </div> <div> <div>📌</div> <div>Determinar esquema de implementación</div> </div>	<div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
9	<div> <div>▼</div> <div> <div>📋</div> <div>Planificar monitoreo y mantenimiento</div> </div> </div> <div> <div>📌</div> <div>Plan de actualización y mantenimiento</div> </div> <div> <div>📌</div> <div>Determinar e implementar acciones del monitoreo</div> </div>	<div>● Done</div> <div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
10	<div> <div>▼</div> <div> <div>📋</div> <div>Producir el reporte final</div> </div> </div> <div> <div>📌</div> <div>Construcción del reporte final</div> </div> <div> <div>📌</div> <div>Presentación del reporte</div> </div>	<div>● New</div> <div>● In Progress</div> <div>● In Progress</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	
11	<div> <div>▼</div> <div> <div>📋</div> <div>Revisar el proyecto</div> </div> </div> <div> <div>📌</div> <div>Documentación de conocimiento adquirido</div> </div>	<div>● Done</div> <div>● Done</div>	<div>Ingrid Rodriguez Nova</div> <div>Ingrid Rodriguez Nova</div>	