

1. Descripción general

Actualmente con la ayuda de las plataformas tecnológicas como Uber, Airbnb o Wallapop es posible generar un ingreso extra para complementar o hasta cubrir los gastos diarios de una persona o familia. Posiblemente tengas una habitación o departamento libre, en caso de que hayas cambiado de residencia o si tus hijos ya se hayan mudado por su cuenta.

Por ejemplo, Carina vive en la colonia Hipódromo en la Ciudad de México. Carina está jubilada y tiene una habitación amueblada que le gustaría rentar para generar un ingreso adicional a su pensión. Su departamento está en una muy buena ubicación, en la calle Cholula, cerca de la colonia Condensa, el Parque España y el Bosque de Chapultepec. Carina está contemplando ofrecer su habitación, o inclusive todo el departamento, en la plataforma de Airbnb. El departamento de Carina se encuentra en excelente estado y cuenta con 65 m² construidos, 2 recámaras con camas matrimoniales, un baño, sala, comedor, cocina y área de lavado. Además, cuenta con servicio Wifi, pero no tiene estacionamiento.

Carina, así como ustedes, obtuvo de la empresa Airbnb una base de datos con la información pública de las ofertas que tiene en cada ciudad del mundo. Con base en dicha información, ¿Cómo puede usar esta información para estimar el precio de renta de la habitación o del departamento? ¿Puedes identificar los factores que le dan plusvalía a una propiedad? ¿Puede aplicarse a propiedades en otras ubicaciones, o inclusive otras ciudades? De manera conjunta, y pensando en un usuario de Airbnb, ¿Podría tener una categorización, donde, de acuerdo con sus intereses, se clasifiquen mobiliarios como recomendables o no recomendables?

2. Declaración del reto

En 5 semanas desarrolla un análisis de datos para determinar el precio de renta de una habitación o departamento en una ciudad del mundo y definir bajo su criterio cuáles propiedades son o no recomendables rentar. De manera independiente y autosuficiente identifica patrones e interacciones relevantes de la información que provee la plataforma tecnológica Airbnb.

3. Datos

Puedes acceder al sitio Inside Airbnb para descargar el archivo de los anuncios de Airbnb en la Ciudad Seleccionada. Descarga el archivo listings.csv.gz que contiene la información más detallada de los anuncios en renta.

4. Producto analítico

Identifique un usuario final que puede estar interesado en un producto de analítica basado en estos datos y diseñe su producto pensando especialmente en ese usuario. Esto

quiere decir que su desarrollo debe ser completo, pero debe estar especialmente enfocado en los intereses de este usuario. Su producto debe estar en forma de un **tablero analítico** que contenga elementos de visualización y los modelos seleccionados para responder a las preguntas de negocio. Los modelos deben estar serializados y deben poder emplearse de manera interactiva en el tablero. El tablero debe ser desplegado en la nube empleando contenedores. El nivel esperado de desarrollo de este producto es de **prototipo funcional**.

5. Equipo y Roles

El equipo estará conformado por 2 estudiantes del Tec y 2 de Uniandes. Para la realización de este proyecto se han contemplado los siguientes roles:

1. Ingeniería de datos. (Tec).
2. Análisis de datos. (Tec).
3. Ciencia de datos regresión. (Tec, Andes).
4. Ciencia de datos clasificación. (Tec, Andes).
5. Análisis de negocio. (Tec, Andes).
6. Tablero de datos. (Andes).
7. Despliegue y mantenimiento. (Andes).

Todos los miembros del equipo deben participar en el rol de Análisis de Negocio. Excluyendo el rol de Análisis de Negocio, cada miembro debe seleccionar dos roles más a desarrollar. Ejemplo de distribución de roles:

- Estudiante 1 Tec: Análisis de Negocio + Ingeniería de datos + Análisis de datos
- Estudiante 2 Tec: Análisis de Negocio + Ciencia de datos regresión + Ciencia de datos clasificación
- Estudiante 1 Andes: Análisis de Negocio + Tablero + Despliegue
- Estudiante 2 Andes: Análisis de Negocio + Ciencia de datos regresión + Ciencia de datos clasificación

Otro ejemplo de distribución de roles:

- Estudiante 1 Tec: Análisis de Negocio + Ingeniería de datos + Ciencia de datos regresión
- Estudiante 2 Tec: Análisis de Negocio + Análisis de datos + Ciencia de datos clasificación
- Estudiante 1 Andes: Análisis de Negocio + Tablero + Ciencia de datos regresión

- Estudiante 2 Andes: Análisis de Negocio + Despliegue + Ciencia de datos clasificación

La calificación de cada miembro del equipo estará asociada a las tareas específicas de los roles tomados y al resultado global del proyecto. **Nota para estudiantes de Uniandes: los roles de cada miembro del equipo (diferentes a análisis de negocio) deben ser diferentes a los seleccionados en el proyecto 1.**

Como **canal de comunicación** tendremos Slack, sin embargo, los integrantes del grupo pueden encontrar otros espacios para conectarse y comunicarse. LA cuenta de Slack es la siguiente: https://join.slack.com/t/1236tecuniandes/shared_invite/zt-3fquiemtc-h_Y8TeEckrl5qlKhy2ZY2w

6. Etapas y entregables

Rompehielo

Video: Debes hacer un video de presentación personal (2-4 minutos) donde nos cuentes sobre ti, quién eres, de dónde eres, qué estás estudiando, qué semestre estás, cuáles son tus hobbies, áreas de interés, planes futuros. El estilo es libre, se creativo.

Comenta a tus compañeros: Después de haber publicado tu video, debes comentar al menos dos videos de estudiantes de la otra universidad y mínimo uno de un compañero de clase, con lo que más te interesa de su presentación, si comparten gustos, intereses o curiosidades.

Enlace: Para desarrollar esta actividad, usaremos el siguiente Padlet:

<https://globalclassroom.padlet.org/global/1326-tec-uniandes-rompehielo-26oys8qijof>

contraseña: global1



Etapa 1: Planeación del proyecto de análisis de datos

Roles involucrados: **Análisis de negocio**

Entregables

Project charter: documento de máximo 1 página que debe contener

1. Descripción de la situación problemática
2. Preguntas de negocio a resolver
3. Objetivos del proyecto (utilizar el método SMART)
4. Identificación de limitaciones y restricciones
5. Plan de trabajo con cronograma y responsabilidades
6. Referencias bibliográficas

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 2 - Extracción, limpieza y transformación de datos

Roles involucrados: Ingeniería de datos.

Cargue el subconjunto de datos en python, explore los datos disponibles y realice una limpieza cuidadosa. Identifique datos faltantes y decida un estrategia para su gestión. Asegúrese de que los datos queden en un formato que permita su posterior análisis. Genere nuevas características de ser necesario. Documente los procedimientos de extracción, limpieza y alistamiento realizados.

Entregables

Sección de máximo 1 página que debe contener una descripción del conjunto de datos inicial, los hallazgos y procedimientos de limpieza realizados, y las características del conjunto de datos final.

Cuaderno de Jupyter donde se presenten en detalle todos los pasos de limpieza de los datos obtenidos de Airbnb de la Ciudad Seleccionada. El Notebook debe considerar al menos los siguientes requisitos:

1. Análisis exploratorio de los datos iniciales (Número de variables, tipo, datos nulos).
2. Proceso de preparación, limpieza y transformación de datos.
3. No hay datos faltantes o "NA".
4. Todas las variables son congruentes en su formato (fecha, hora, texto).
5. Las variables categóricas consistentes en formato.
6. Variables categóricas y numéricas codificadas.
7. Solo variables potencialmente relevantes para el análisis posterior.
8. Análisis y tratamiento de outliers.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 3 - Análisis Exploratorio de Datos

Roles involucrados: Análisis de datos.

Realice un análisis exploratorio que permita describir estadística y visualmente el comportamiento de las variables a considerar. Calcule estadísticas descriptivas, realice histogramas, diagramas de caja, diagramas de dispersión, diagramas de violín y otros que permitan comprender cómo se comportan las variables. Genere nuevas características de ser necesario. Documente el análisis realizado.

Entregables

Sección de máximo 1 página que debe contener una descripción de los hallazgos principales del análisis exploratorio y su relación con las preguntas de negocio.

Cuaderno de Jupyter donde se presente en detalle el análisis exploratorio realizado.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 4 - Modelos predictivos

Tras explorar en detenimiento los datos y tener claras la preguntas de negocio, es hora de pasar a construir los modelos. Se espera que sean modelos de regresión y clasificación. Los integrantes del Tec deben desarrollar modelos estadísticos y los integrantes de Unianandes modelos de redes neuronales. Al desarrollar los modelos, tenga presente lo aprendido en la exploración de datos, así como el usuario final seleccionado y las preguntas a resolver:

1. Dado un determinado departamento en la ciudad seleccionada, ¿cuál será el precio más adecuado por noche?
2. Dadas las condiciones de la BD, ¿que consejos les podemos dar a los anfitriones para incrementarse el precio de las rentas por Airbnb?
3. ¿Cómo clasificaríamos los departamentos en recomendados o no recomendados?

Etapa 4a - Modelos predictivos Tec

Roles involucrados: Ciencia de datos regresión (Tec), Ciencia de datos clasificación (Tec).

Mediante modelos de regresión avanzados, genere un modelo inicial, revisando su desempeño y haciendo cambios para optimizar su poder predictivo que permita contestar las preguntas anteriores.

Entregables

Sección de máximo 1 página que debe contener por lo menos:

1. Modelo inicial de regresión.
2. Análisis numéricos y gráficos de los residuales.
3. Análisis de los índices de correlación y potencia.
4. Análisis de normalidad de los datos.
5. Optimización del modelo. Descripción e interpretación.

Cuaderno de Jupyter donde se presente en detalle el proceso de modelado y sus resultados.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 4b - Modelos predictivos Andes

Roles involucrados: **Ciencia de datos regresión** (Andes), **Ciencia de datos clasificación** (Andes).

Mediante modelos de redes neuronales, genere modelos iniciales, revisando su desempeño y haciendo cambios para optimizar su poder predictivo para contestar las preguntas anteriores.

Entregables

Sección de máximo 1 página que debe contener por lo menos:

1. Modelos iniciales de regresión y clasificación.
2. Ingeniería de características (si se realiza).
3. Búsqueda de amplia de hiperparámetros.
4. Análisis de resultados.

Cuaderno de Jupyter donde se presente en detalle el proceso de modelado y sus resultados.

Soportes de los experimentos realizados en MLflow.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 5 - Diseño y desarrollo del tablero

Roles involucrados: Tablero de datos.

Tras explorar los datos y construir los modelos, es hora de diseñar y desarrollar el producto. El producto debe ser un tablero en Dash desplegado en la nube de AWS, usando **contenedores Docker**. El tablero debe ser de fácil uso y le debe permitir al usuario acceder a **3 visualizaciones** relevantes y emplear los el modelos predictivos ingresando los datos apropiados.

Empiece por diseñar el tablero: ¿qué valores debe permitir ingresar? ¿qué resultados genera? ¿qué visualizaciones incluye? ¿cómo mostrará las instrucciones? ¿cómo dispondrá estos elementos en el tablero? Para esta tarea es buena idea hacer un wireframe (un diseño sencillo que puede hacer en papel o digitalmente), que le permite tener una visión clara de su tablero y todos sus elementos. Recuerde no perder de vista al usuario y su necesidad. Piense siempre en la experiencia del usuario. Una vez haya terminado el diseño, desarrolle su tablero en Dash.

Entregables

Sección de máximo 1 página que debe contener por lo menos:

1. Diseño inicial del tablero (mockup).
2. Descripción del tablero, su funcionalidad y uso.
3. Pantallazos del tablero desarrollado que apoyen la descripción.

Fuentes del tablero (archivos .py del tablero, assets, archivos html, css y otros de estilo.).

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 6 - Despliegue y mantenimiento

Roles involucrados: Despliegue y mantenimiento.

El tablero debe quedar desplegado en la nube empleando contenedores Docker en AWS. Además, los **modelos** que se desplieguen deben estar **serializados** en archivos. Es decir, al ejecutar su tablero los modelos se deben cargar pre-entrenados de archivos locales. Asegúrese de que su tablero sea accesible y quede en ejecución.

Entregables

Sección de máximo 1 página que debe contener por lo menos:

1. Descripción del proceso de despliegue con instrucciones para lanzar el tablero a partir del código disponible en el repositorio de Github.
2. Pantallazos del despliegue que apoyen la descripción. Asegúrese de que sean visibles las direcciones privada y pública de la máquina que usa para lanzar los contenedores, así como su usuario en AWS.
3. Soportes del tablero desplegado.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 7 - Evaluación

Roles involucrados: ~~Análisis de datos~~.

Evalúe los resultados, considerando las preguntas de negocio, los *insights* obtenidos del análisis descriptivo, los modelos predictivos y el tablero desarrollado.

Entregables

Sección de máximo 1 página que debe contener por lo menos:

1. Respuesta a cada una de las preguntas de negocio.
2. Relación de cómo las visualizaciones, análisis y modelos responden a cada una de las preguntas de negocio.
3. Lecciones aprendidas.

Fecha de entrega: lunes 1 de diciembre de 2025

Etapa 8 - Presentación

Roles involucrados: **todos.**

Como paso final del proyecto, los miembros del equipo realizarán una presentación en la que presentarán sus hallazgos principales y demostrarán el producto analítico. Cada miembro debe presentar los resultados asociados a su rol. Al momento de la presentación el tablero debe estar desplegado y listo para ser usado públicamente. La presentación tiene una duración máxima de **10 minutos** (estrictos), seguida de una sesión de preguntas. Se ofrecerán espacios para la presentación y cada equipo podrá seleccionar uno.

Entregables

Archivo o enlace a las dispositivas empleadas en la presentación.

Fecha de entrega: lunes 1 de diciembre de 2025

Fecha de presentación: jueves 4 de diciembre de 2025

Reflexión Final

Paso 1: En el primer parte, los estudiantes **de manera individual van a autoevaluarse** y evaluar el rendimiento de sus compañeros en las diferentes entregas del proyecto. En este espacio, con un google forms, se les preguntará sobre su percepción individual de la actividad, que les gusto, que mejorarían.

Paso 2: En Padlet se hará una especie de infografía grupal, donde mediante videos, imágenes, frases, los estudiantes compartirán sus **reflexiones finales** respecto a la actividad. Pueden compartir información respecto a cómo se sintieron, que aprendieron de los otros países y culturas, qué les gustó y qué no, de la actividad realizada, cómo la mejorarían.

padlet reflexión:

<https://globalclassroom.padlet.org/global/1326-tec-uniandes-reflexi-n-q411wjfr6936y>
contraseña: global2

Repositorio Git

Todas las tareas realizadas por los integrantes de los Andes (excepto las asociadas al análisis de negocio) deben estar soportadas en commits realizados en el repositorio Git. El repositorio debe contener una carpeta para cada etapa del proyecto con los soportes asociados a cada tarea. El historial de commits debe reflejar el aporte de cada miembro del grupo (de acuerdo con sus roles). La carpeta asociada a la etapa de despliegue debe incluir todos los archivos necesarios para lanzar el tablero en un contenedor siguiendo las instrucciones incluidas en el reporte.

Sobre los soportes y la calificación

Nota: los soportes son parte fundamental de la entrega. Su no entrega lleva a una alta penalización.

Nota 2: si bien el trabajo es en equipo, la nota es individual, luego es necesario que cada miembro del equipo demuestre su contribución al proyecto, tanto a través de los **commits en el repositorio**, como a través del **reporte** de trabajo en equipo y la **sustentación**.

La calificación individual del proyecto se realizará de la siguiente manera:

1. **10 puntos:** contribución individual a la actividad rompehielo.
2. **20 puntos:** contribución individual al reporte.
3. **20 puntos:** contribución individual a la presentación.
4. **40 puntos:** contribución individual de acuerdo con su rol, reflejado en los entregables y soportes asociados.
5. **10 puntos:** reflexión individual.

Recomendaciones

1. El objetivo del proyecto es lograr un buen producto, bien soportado y claramente desarrollado. Justifique adecuadamente sus decisiones, observaciones y conclusiones.
2. Sea conciso y eficiente con el espacio. Ni el reporte ni la presentación deben ser largos. Al contrario, en un buen reporte cada gráfica y afirmación importa, y en un buena presentación cada diapositiva cuenta.
3. Es un trabajo en equipo. Defina los ítems de trabajo, asígnelos entre los miembros del equipo, defina fechas de entrega y revisión interna. Discuta los resultados, observaciones y conclusiones. Priorice tareas y resultados a incluir.
4. Empiece a trabajar prontamente y discuta con el instructor su avance y resultados.

Fecha de entrega final: lunes 1 de diciembre de 2025