

1. Descripción general

El objetivo de este proyecto es desarrollar un producto de analítica de datos sobre el costo de arrendamiento de apartamentos. Para desarrollar este producto usted debe hacer uso de los datos adjuntos al enunciado del proyecto. 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. El nivel esperado de desarrollo de este producto es de **prototipo funcional**.

2. Roles

Para la realización de este proyecto se han contemplado los siguientes roles:

1. Análisis de negocio.
2. Ingeniería de datos.
3. Análisis de datos.
4. Ciencia de datos.
5. Tablero de datos.
6. Despliegue y mantenimiento.

Cada miembro del grupo debe seleccionar 2 roles, y realizar las tareas asociadas a estos 2 roles. La calificación de cada miembro del equipo estará asociada a las tareas específicas de los roles tomados y al resultado global del proyecto. Estos roles se intercambiarán en cada proyecto, es decir, cada miembro del equipo debe tomar un rol una vez en el semestre.

División sugerida de roles:

1. Análisis de negocio + Tablero de datos.
2. Ingeniería de datos + Análisis de datos.
3. Ciencia de datos + Despliegue.

3. Comprensión de negocio y datos: preguntas de negocio y plan de acción

Tarea 1 - Preguntas de negocio

Determine la pregunta (o preguntas) de negocio que quiere resolver para su cliente seleccionado.

Identifique cómo puede resolver estas preguntas a través de visualizaciones de los datos (descriptivo) y un modelo predictivo (preferiblemente de regresión, pero también puede ser de clasificación). Seleccione buenas **preguntas**, relevantes para su cliente y asegúrese de **responder** a ellas con los **resultados** del proyecto.

Roles involucrados: Análisis de negocio, Análisis de datos, Ciencia de datos.

Lidera (responsable): Análisis de negocio.

Tarea 2 - Exploración 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. Documente el análisis realizado.

Roles involucrados: Análisis de datos.

4. Datos

Tarea 3 - Preparación de los datos

Explore los datos disponibles y realice una limpieza cuidadosa. Identifique datos faltantes y decida una estrategia para su gestión. Trate los datos categóricos apropiadamente, considerando que serán usados posteriormente en el modelamiento. Genere nuevas características de ser necesario. Asegúrese de que los datos queden en un formato que permita su posterior análisis. Documente los procedimientos realizados.

Roles involucrados: Ingeniería de datos.

5. Modelos

Tras explorar en detenimiento los datos y tener claras las preguntas de negocio, es hora de pasar a construir los modelos. Se espera que sean modelos de regresión. Tenga presente lo aprendido en la exploración de datos, así como el usuario final seleccionado y las preguntas a resolver.

Tarea 4 - Modelamiento

Aquí deberá explorar diferentes configuraciones de modelo, realizar ingeniería de características, emplear diferentes métodos de estimación, comparar y seleccionar las mejores alternativas. Consulte bibliografía que le permita contar con elementos para proponer los modelos. No es necesario emplear todas las variables disponibles, pero todas las variables incluidas y sus relaciones deben estar correctamente justificadas. Evalúe sus modelos usando métricas apropiadas (diagnóstico, ajuste, información). Documente el modelamiento realizado.

Roles involucrados: Ciencia de datos.

6. Producto y evaluación

Tras explorar los datos y construir los modelos, es hora de diseñar y desarrollar el producto. El producto debe ser un tablero de datos en Dash. El tablero debe ser de fácil uso y le debe permitir al usuario acceder a visualizaciones relevantes y emplear el modelo predictivo ingresando los datos apropiados.

Tarea 5 - Diseño y desarrollo del tablero

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.

Roles involucrados: Tablero de datos.

Tarea 6 - Evaluación

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

Roles involucrados: Análisis de negocio, Análisis de datos, Ciencia de datos, Tablero de datos.

Lidera: Análisis de negocio.

Tarea 7 - Despliegue y mantenimiento

Despliegue el tablero en una instancia (máquina virtual) en EC2. Asegúrese de que su tablero sea accesible y quede en ejecución. Su tablero deberá estar disponible al momento de la sustentación y su operación se deberá demostrar en la misma.

Roles involucrados: Despliegue y mantenimiento.

7. Entregables

Como resultado de las tareas anteriores deberá entregar los siguientes resultados y soportes:

1. Resultado 1: reporte de **máximo 7 páginas** con la documentación de las tareas, los resultados principales del análisis exploratorio de datos y la modelización. Cada sección del proyecto debe indicar qué rol y miembro del equipo la realizó. El reporte debe describir cada tarea listada en el reporte.

2. Resultado 2: presentación de **máximo 10 minutos** con los resultados principales del análisis exploratorio de datos y la modelización. Esta presentación debe incluir también un espacio para demostrar el tablero desarrollado. Cada miembro del equipo debe demostrar sus resultados (asociados a sus roles dentro del proyecto).
3. Resultado 3: tablero desarrollado en Dash y desplegado en la nube en un ambiente de pruebas.
4. Reporte de trabajo en equipo: incluya un pequeño reporte de cómo se dividieron los roles entre los miembros del equipo.
5. Soporte 1: análisis de negocio (soporte incluido en el reporte).
6. Soporte 2: fuentes de limpieza (cuadernos de jupyter o archivos .py con la limpieza de datos).
7. Soporte 3: fuentes de análisis (cuadernos de jupyter o archivos .py con el análisis exploratorio).
8. Soporte 4: fuentes de modelización (cuadernos de jupyter o archivos .py con la modelización desarrollada). Aquí debe incluir los pasos de entrenamiento, prueba y evaluación del modelo.
9. Soporte 5: fuentes del tablero (archivos .py del tablero desarrollado).
10. Soporte 6: snapshots de la máquina lanzada (de EC2 y terminal), y URL del tablero en ejecución.
11. Soporte 7: **repositorio Git** en Github **público**, con un historial de commits que claramente refleje el aporte de cada miembro del grupo (de acuerdo con su rol). El repositorio debe estar estructurado con carpetas que reflejen las tareas definidas en el proyecto. Debe incluir una carpeta “despliegue” que contenga la última versión del tablero y permita lanzarlo, replicando el despliegue en AWS.

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 (de 3 personas), 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. **25 puntos:** contribución individual al reporte.
2. **25 puntos:** contribución individual a la presentación.
3. **50 puntos:** contribución individual de acuerdo con su rol, reflejado en los entregables y soportes asociados.
4. **3 puntos:** descripción de su aporte en el reporte de trabajo en grupo.

8. 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 una 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: lunes 24 de febrero, 23:59 p.m.