

1. Descripción general

El objetivo de este proyecto es desarrollar un producto de analítica de datos sobre el uso de servicios de bicicletas compartidas. Se han identificado dos usuarios finales interesados en este producto:

1. El área de operaciones del operador del sistema, que debe asegurar que la operación atienda la demanda con un buen nivel de servicio a lo largo del año. Para esto debe contratar personal, tener disponibilidad de vehículos de transporte de bicicletas y equipos de mantenimiento de las mismas. La Secretaría de Movilidad de la ciudad, interesada especialmente en el nivel de uso de las bicicletas, su costo para las finanzas públicas y la operación que debe realizarse para .
2. El área de finanzas del operador del sistema, que debe asegurar que los ingresos recibidos sean suficientes para cubrir todos los costos (operación, mantenimiento, capital) y alcance la rentabilidad esperada.

Seleccione uno de los dos usuarios finales 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. Ingeniería de datos.
2. Análisis de datos.
3. Ciencia de datos.
4. Análisis de negocio.
5. Tablero de datos.
6. Despliegue.

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 + Ciencia de datos.
3. Análisis de datos + Despliegue.

3. Preguntas de negocio y plan de acción

Tarea 1

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.

4. Datos

Para desarrollar este producto usted debe hacer uso de los datos adjuntos al enunciado del proyecto.

Tarea 2 - Limpieza y alistamiento de datos

Cargue los datos en python, 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 para análisis y modelamiento. Asegúrese de que los datos queden en un formato que permita su posterior análisis. Documente los procedimientos realizados.

Roles involucrados: Ingeniería de datos.

Tarea 3 - 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.

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, aunque también pueden ser de clasificación. 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. Como hay un buen número de variables, se espera que el modelo desarrollado por cada equipo sea único. Evalúe sus modelos usando métricas apropiadas (diagnóstico, ajuste, información). Documente el modelamiento realizado e incluya al menos 2 referencias como justificación.

Roles involucrados: Ciencia de datos.

6. Producto

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, usando una máquina virtual. 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 - Despliegue

Despliegue el tablero en una instancia (máquina virtual) en EC2. Asegúrese de que su tablero sea accesible y quede en ejecución.

Roles involucrados: Despliegue.

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.

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.
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, 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 10 de septiembre, 2:00 p.m.