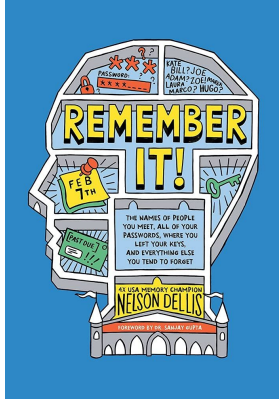




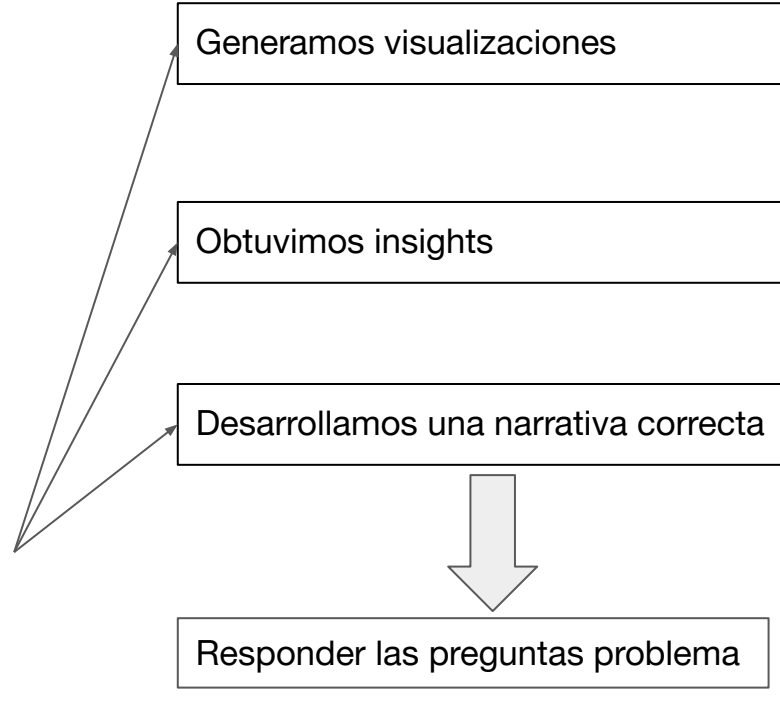
Obtención de insights a partir de visualizaciones

Deberán entregar en duplas de dos estudiantes la segunda pre entrega de su Proyecto Final. Entrenarán y optimizarán versos modelos de machine learning para resolver una problemática específica, detectada en la instancia de entrega anterior. El objetivo es que puedan utilizar modelos de machine learning para resolver el problema de una industria o negocio.

Recordemos...



Clase 14
Desafío entregable:
Obtención de insights



CODER HOUSE



Obtención de insights a partir de visualizaciones

Objetivos generales

- ✓ Obtener datos de diversas fuentes como APIs o Bases de datos públicas para luego analizarlos mediante el lenguaje Python con el fin de contestar una pregunta de interés para una industria, negocio o proyecto personal. Se deberán utilizar datasets complejos implementando técnicas avanzadas para la limpieza y adquisición de datos

Objetivos específicos

- ✓ Estructurar un problema en función de múltiples pero simples preguntas/hipótesis a responder
- ✓ Importar datos crudos de APIs o bases de datos usando Python
- ✓ Limpiar y transformar los datos para permitir un posterior análisis
- ✓ Contar una historia mediante el análisis exploratorio de datos



Obtención de insights a partir de visualizaciones

Requisitos base

- ✓ Un notebook (Colab o Jupyter) que debe contener:
 1. **Abstracto con motivación y audiencia:** Descripción de alto nivel de lo que motiva a analizar los datos elegidos y que audiencia se podrá beneficiar de este análisis
 2. **Preguntas/hipótesis que queremos responder:** Lista de preguntas que se busca responder mediante el análisis de datos. Bloques de código donde se importan los datos desde una API o base de datos pública y los guarda en un archivo local csv o json. El estudiante puede luego de descargar los datos, comentar este bloque de código
 3. **Análisis exploratorio de datos (EDA):** Análisis descriptivo de los datos mediante visualizaciones y herramientas estadísticas



Obtención de insights a partir de visualizaciones

Requisitos base

- ✓ Una presentación (PDF; PowerPoint o Google Slides) que debe contener
 1. **Abstracto con motivación y audiencia:** Descripción de alto nivel de lo que motiva a analizar los datos elegidos y que audiencia se podrá beneficiar de este análisis
 2. **Resumen de metadata:** resumen de los datos a ser analizados es decir, número de filas/columnas, tipos de variables, etc
 3. **Preguntas hipótesis que queremos responder:** Lista de preguntas que se busca responder mediante el análisis de datos
 4. **Visualizaciones ejecutivas que responden nuestras preguntas:** utilización de gráficos que responden las preguntas de interés de nuestro proyecto.
 5. **Insights:** resumen de hallazgos del proyecto. Aquí consolidamos las respuestas a las preguntas/hipótesis que fuimos contestando con las visualizaciones



Obtención de insights a partir de visualizaciones

Sugerencias

Es conveniente retomar el dataset trabajado en la primera pre entrega y enriquecerlo (e.g joins, y creación de nuevas columnas) con información proveniente de APIs públicas siempre que se pueda con el fin de practicar las nuevas habilidades adquiridas. Se recomienda retomar la metodología de trabajo y reutilizar algoritmos ya entrenados, de ser necesario.

Requisitos extra

- ✓ Subir el proyecto a Github



Obtención de insights a partir de visualizaciones

Dont's

- ✓ Utilizar jerga demasiado técnica en la presentación (recordar que la audiencia de la misma son roles ejecutivos)
- ✓ Sobrecargar las diapositivas
- ✓ Realizar una presentación con más de 12 slides de extensión

Modelo de Proyecto final

- ✓ [Proyecto final \(Notebook\)](#) (Se debe abrir con Google Collaboratory o Jupyter Notebook)
- ✓ [Ejemplo Presentación](#)

Glosario

Revisión de pares: proceso constructivo donde se busca mejorar la calidad de un proyecto de ciencia de datos. Funciona como un mecanismo de control antes de presentar los resultados obtenidos

Narrativa de presentación: se refiere a la metodología que se utiliza para transmitir información y resultados ante una audiencia. Se recomienda ser lo más simple posible a la hora de presentar resultados y utilizar jerarquías.

Reglas para presentación de resultados: se refiere a 4 puntos importantes a la hora de mostrar resultados a una audiencia. Debemos tener cuidado del texto, tablas, gráficos y las estadísticas que se reportan para asegurar que el estudio mantenga un buen estándar de calidad.