

Tarea 1

EDA y modelos bayesianos

Integrantes: Matias Arriola
Leopoldo Cárdenas
Francisco Muñoz
Gabriel Ortega
Profesor: Nicolás Caro
Auxiliar: Rodrigo Lara

Fecha de realización: 09 de mayo de 2020
Fecha de entrega: 17 de mayo de 2020
Santiago, Chile

Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Índice de Contenidos

1. Carga y limpieza de datos	1
2. EDA	1
3. Regresión Lineal Bayesiana	1

P1) Carga y limpieza de datos

Para iniciar con la carga y limpieza de datos, primero se creo el módulo “paths” en donde se hará registro de todas las rutas necesarias para realizar la tarea. Además, se generan las rutas `PATH_DATA` y `PATH_RAW`, que serán a las carpetas "data" y "raw" respectivamente. Luego se utilizará el diccionario `dict_path_raw`, para que al entregarle la llave "wNN", indicará la ruta de la carpeta "wNN" en "raw". Sumado a este diccionario se crean otros dos, `dict_csv_mc_a` y `dict_csv_mc_f`, con el fin de que entreguen la ruta para los archivos .csv, "metrocuadrado_all_wNN" y "metrocuadrado_furnished_wNN" respectivamente, al darles de llave "wNN".

Después, con el fin de generar "furnished" de eliminar los datos duplicados, se procede a crear un único DataFrame que contenga la información de los datasets proporcionados. Primero se crean dos listas de DataFrames, uno para "all" el otro para "furnished", para luego concatenar los DataFrames en dos DataFrames únicos, nuevamente uno para "all" el otro para "furnished". Así, se crea una nueva columna "furnished", que informa si el DataFrame proviene de "furnished", que será indicado con un 1, o si viene de "all", indicado con un 0. Posteriormente se unen los dos DataFrames anteriores en uno único, añadiendo la columna extra creada, para finalmente eliminar los duplicados con la función `drop_duplicates` de pandas.

P2) EDA

Hola soy la pregunta 2

P3) Regresión Lineal Bayesiana