Proyecto: Análisis Exploratorio de los Datos completo.

instituto tecnológico de morelia

Proyecto.

M.I.A. WILFRIDO CORTES OROZCO

2023

|  |  |  |
| --- | --- | --- |
| Nombre del Alumno | Número de control | Número de Proyecto |
| Brenda Cortes Aguilar | 20120097 | 1 |

# Descripción y objetivos del proyecto.

Objetivos:

* Que el estudiante pueda utilizar Git para llevar a cabo el control de versiones.
* Que el estudiante pueda realizar el análisis exploratorio de los datos completo sobre un dataset real.
* Que el estudiante ponga en práctica sus conocimientos teóricos y ejercite su capacidad de análisis.
* Que el estudiante genere un dataframe listo para usarse en el entrenamiento de un modelo de ML.

Descripción de la actividad:

Este proyecto consiste en tres etapas principales:

1. Análisis y exploración de los datos.
2. Tratamiento de los datos.
3. Manipulación y preparación de los datos.

En cada etapa se van a realizar varios pasos para ir explorando y preparando el set de datos.

Las instrucciones vienen en el archivo de Jupyter-Notebook, el trabajo debe ser realizado sobre ese archivo y debe ser subido a GitHub en un repositorio nuevo, el repositorio debe ser llamado “House Pricing EDA - NombreEstudiante”, es importante llevar ese formato ya que, posteriormente en el curso, se utilizará para mostrar en el portafolio de proyectos.

Cada repositorio creado debe tener un archivo “README” con la información del archivo, así como el archivo Jupyter-Notebook, el dataset a utilizar y este archivo de instrucciones. Dentro del README se debe detallar en qué consiste el análisis, qué pasos se siguen, cómo se hacen las conclusiones y qué es lo que se genera al final (un dataframe numérico listo para su uso).

Como sugerencia, hay que ir tomando capturas de las gráficas, de los procesos más importantes y de las conclusiones, serán útiles para más adelante, hay que guardarlas en el mismo repositorio en una carpeta llamada “Screenshots” o similar.

Para entregar el proyecto, se deben enviar los siguientes archivos/datos:

* El archivo de Jupyter, el dataset y todo lo necesario para ejecutarlo.
* Este archivo de instrucciones con el nombre del alumno, número de control y número de proyecto.
* El enlace al repositorio público en GitHub del proyecto.

Métrica de evaluación:

El proyecto tiene un valor de 30 puntos sobre la calificación final del curso, el puntaje estará dado por:

|  |  |  |
| --- | --- | --- |
| Criterio | Valor | Métrica |
| Se siguen los pasos correctamente. | 15 puntos. | Excelente: 15 puntos (todos los puntos realizados correctamente).  Bien: 13 puntos (máximo 2 puntos no realizados totalmente bien o incompletos).  Regular: 10 puntos (hasta 5 puntos incompletos o con errores).  Apenas aceptable: 8 puntos (máximo 7 puntos no realizados correctamente, inconsistentes o con errores).  No aceptable: 0 puntos (más de 7 puntos incorrectos, incompletos o con errores), anulación del proyecto. |
| El dataframe resultante es numérico y está listo para usarse. | 5 puntos. | Correcto: 5 puntos. El dataframe generado es totalmente numérico y se tienen identificadas las variables cuya correlación con la variable objetivo es más fuerte (10 primeros).  Incorrecto: 0 puntos. El dataframe generado no es totalmente numérico y/o no se tienen identificadas las variables más correlacionadas con la variable objetivo. |
| Análisis y gráficas claras. | 5 puntos. | Adecuado: 5 puntos. Las gráficas son seleccionadas correctamente para el tipo de dato, los análisis sobre cada paso son correctos/coherentes.  Irregular: 2 puntos. Las gráficas seleccionadas no son las adecuadas para el tipo de dato seleccionado pero los análisis sí son coherentes.  Inadecuado: 0 puntos. Las gráficas son totalmente incorrectas para el tipo de dato analizado, los análisis no están presentes o no son coherentes. |
| Conclusiones coherentes y consistentes con el aprendizaje y actividad realizada. | 5 puntos. | Preciso: 5 puntos. Las conclusiones son acertadas y el trabajo presentado cumple con lo necesario para ser considerado como completado.  Inconcluso: 0 puntos. No hay conclusiones o no están correctamente aterrizadas, no se llegó a un análisis adecuado. |