POC(Proof Of Concept)

Summary :

This Project involved in Processing the data like filling the missing values with respective other columns grouping them and filling the mean(Bathroom,Sqrt_ft,Garage and Fireplaces_avg). With HOA and Lot_Acres we had used the Log and Cube root to Normalize the columns and filled the null with the median .And 64 records deleted as they were outliers as per the lot_acers .
For the Kitchen feature and Floor covering :Extracted and analyzed feature frequencies, selecting those present in at least 5% of houses. Missing values were filled using features found in at least 60% of houses,And did One Hot Encoding.
For the Longitude and Latitude:Added a new column for the distance from the center point and two more columns indicating direction, where values were set to 1 if longitude or latitude was greater than the center, otherwise 0.
Finally :
We had 40 columns and 4929 rows.-link

Specific actions and Technologies:

- Google.Colab: Colab is Notebook which is used to execute code on Google Cloud servers
- Google Drive : Used to Store the raw data set and can easily connect to the Google_Colab and also cleaned data set .
- Python and Pandas : Utilized python and Pandas Library to construct a DataFrame for data processing
- Numpy : Utilized for mathematical calculations like (Log,Cube Root)
- Seaborn and Matplotlib: Used for graphical representation to understand data better.
- Plotly.express : Utilized for the graphs ,especially the geo-graphical representation and also few scatter plots -(reference)
- Haversine & Unit: Used to calculate the distance between the two points ,based on Longitude and Latitude.(reference)