Preliminary Proposal: Data Analysis of Real Estate Market in Different Suburbs in South Australia

**Overview**

The objective of this data analysis project is to examine the real estate market in different suburbs of South Australia using the dataset provided by the South Australian Government's Data Portal (<https://data.sa.gov.au/data/dataset/metro-median-house-sales>). The dataset contains information on median house sales prices in various suburbs over a period of time. By analysing this data, we aim to gain insights into the trends and patterns affecting the housing market in different suburbs of South Australia before and after COVID 19.

**Group members:**            Jordan Kane, Van Hoang, Isha Sharma & Magdalene Singh

**Objectives**

1. **Suburb Comparison:** Compare and contrast the median house sales prices in different suburbs to identify high-performing and low-performing areas.
2. **Trends over Time:** Analyze the trends in median house sales prices over time to understand the overall market trajectory in South Australia.
3. **Correlation with Factors:** Explore the potential correlation between house prices and other factors such as location of the suburbs.
4. **Predictive Analysis:** Develop predictive models to forecast future median house sales prices in various suburbs.

**Data Source**

The primary data source for this analysis will be the dataset provided by the South Australian Government's Data Portal (<https://data.sa.gov.au/data/dataset/metro-median-house-sales>). The dataset includes historical records of median house sales prices in different suburbs in South Australia. Potentially using Geoapify APIs to obtain information about the factors (for example, restaurants, shops, ecg) that can affect the

**Methodology**

1. **Data Collection:** The first step will involve collecting the dataset from the provided website. The Data Cleaning Jupyter Notebook includes the steps by steps to combine data from the website, clean the data, and ensure the final data combination is reliable for analysis.
2. **Data Preprocessing:** Perform necessary data preprocessing steps such as handling missing values and cleaning unnecessary data.
3. **Exploratory Data Analysis (EDA):** Conduct exploratory data analysis to gain initial insights into the dataset, visualize trends, and identify outliers or anomalies.
4. **Suburb Comparison:** Compare median house sales prices in different suburbs using statistical measures and visualization techniques. Identify top-performing and underperforming suburbs.
5. **Correlation Analysis:** Investigate potential correlations between house prices and other factors, such as location
6. **Insights:** Summarize the findings from the analysis and provide insights to stakeholders in the real estate industry.

**Research questions:**

1. What are the overall trends in the South Australian real estate market over the past few years? Are house prices increasing or decreasing on average?
2. "How has COVID-19 impacted property pricing trends in South Australia, Or "Has COVID-19 impacted property pricing trends in South Australia?"
3. How do the median house prices vary across different regions or districts within South Australia?

**Tools and Technologies**

* Programming Language: Python
* Data Manipulation: Pandas, NumPy
* Data Visualization: Matplotlib
* Jupyter Notebooks for interactive development and documentation.

**Deliverables**

1. A comprehensive report detailing the findings of the data analysis, including visualizations and insights.
2. Jupyter Notebooks containing the code and step-by-step analysis.
3. Interactive visualizations to assist in presenting the results.
4. A presentation summarizing the key findings and recommendations.

**Conclusion**

This data analysis project aims to provide valuable insights into the real estate market in different suburbs of South Australia. By analyzing historical median house sales prices, we will gain a deeper understanding of the factors influencing property prices and be able to make informed predictions about future market trends. The results of this analysis will be valuable to real estate professionals, investors, and policymakers, aiding them in making strategic decisions based on data-driven evidence.