

Airbnb Dynamic Pricing Recommendation Engine

Introduction

The purpose of this project is to develop a dynamic pricing recommendation system for Airbnb listings. By analyzing historical Airbnb data based on various factors such as city, property type, review scores, and other listing features, the project aims to predict optimal pricing for hosts to maximize revenue while staying competitive. This project highlights the importance of data-driven pricing strategies in the short-term rental market.

Abstract

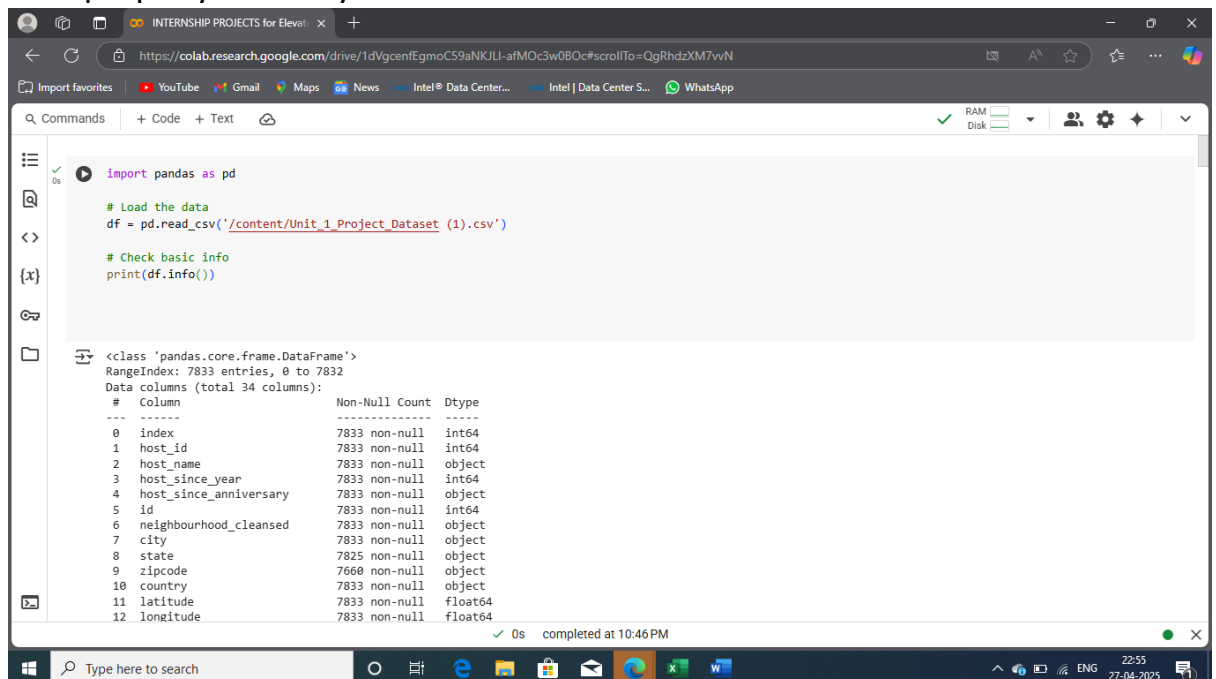
In this project, historical Airbnb listing data was collected and analyzed to identify key pricing predictors. Exploratory data analysis (EDA) was conducted to study price variations across different cities, property types, and quality indicators (e.g., review scores). A regression model was developed to predict suggested prices. Visualizations were created to better understand the data and help users interact with pricing recommendations. The project was implemented using Python, Tableau, and Excel, culminating in a dashboard and price prediction engine.

Tools Used

- **Python** (Pandas, Seaborn, Plotly, Scikit-learn)
 - **Tableau** (Dashboard creation)
 - **Excel** (Data cleaning and initial exploration)
 - **Google Co lab** (Python environment)
-

Steps Involved in Building the Project

1. **Data Collection:** Gathered Airbnb historical data including city, property type, number of reviews, and pricing details.
2. **Data Cleaning:** Handled missing values, removed outliers, and formatted data properly for analysis.



The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: <https://colab.research.google.com/drive/1dVgcenFgmoC59aNKJU-afMOc3w0BOc#scrollTo=QgRhdxM7vN>. The notebook has a file explorer on the left, a command bar, and a code editor. The code in the editor is as follows:

```
import pandas as pd

# Load the data
df = pd.read_csv('/content/Unit_1_Project_Dataset (1).csv')

# Check basic info
print(df.info())
```

The output of the code is displayed below the editor:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7833 entries, 0 to 7832
Data columns (total 34 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   index               7833 non-null  int64
 1   host_id             7833 non-null  int64
 2   host_name           7833 non-null  object
 3   host_since_year     7833 non-null  int64
 4   host_since_anniversary 7833 non-null  object
 5   id                  7833 non-null  int64
 6   neighbourhood_cleansed 7833 non-null  object
 7   city                7833 non-null  object
 8   state              7825 non-null  object
 9   zipcode            7660 non-null  object
10   country            7833 non-null  object
11   latitude            7833 non-null  float64
12   longitude           7833 non-null  float64
```

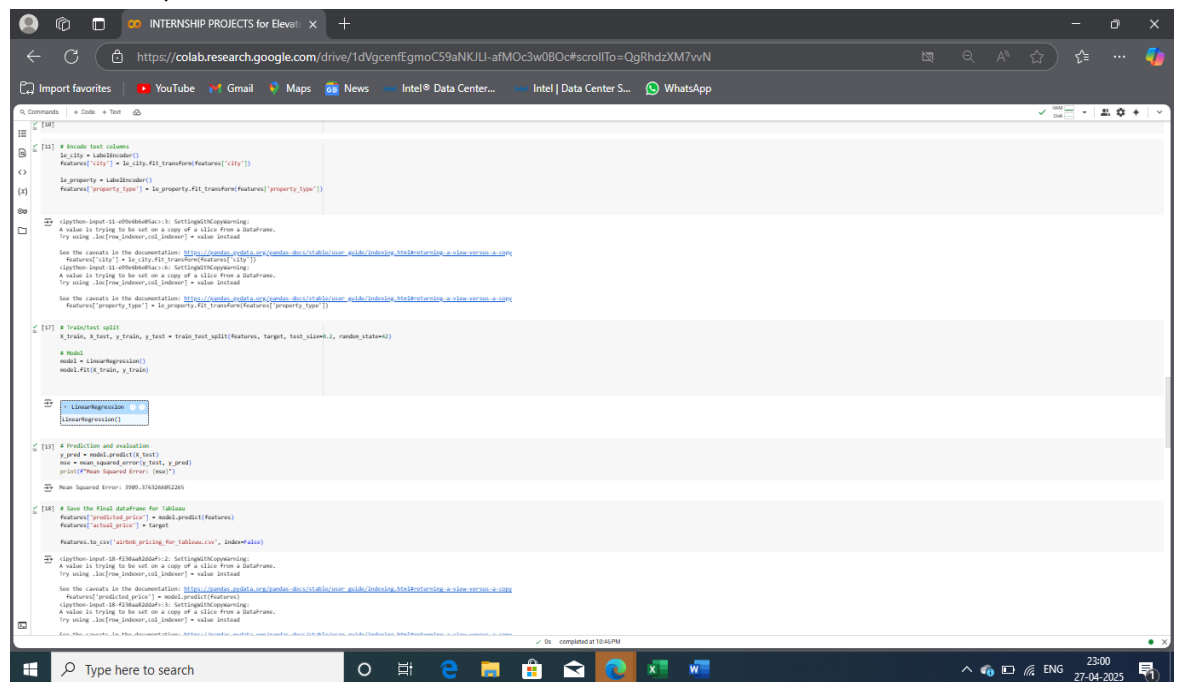
The bottom status bar indicates the code was completed at 10:46 PM.

3. Exploratory Data Analysis (EDA):

- Analyzed average prices by city and property type.
- Studied the impact of review scores on pricing.

4. Model Building:

- Built a Linear Regression model to predict optimal pricing.
- Identified important features like review scores, number of bedrooms, and bathrooms.



```
[18]: # Encode text columns
in_city = LabelEncoder()
Features['city'] = in_city.fit_transform(Features['city'])

in_property = LabelEncoder()
Features['property_type'] = in_property.fit_transform(Features['property_type'])

[19]: # Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(Features, target, test_size=0.2, random_state=0)

[20]: # Create a Linear Regression model
model = LinearRegression()
model.fit(X_train, y_train)

[21]: # Predict and evaluate the model
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error: {}".format(mse))

[22]: # Save the final dataset for Tableau
Features['predicted_price'] = model.predict(Features)
Features['actual_price'] = target

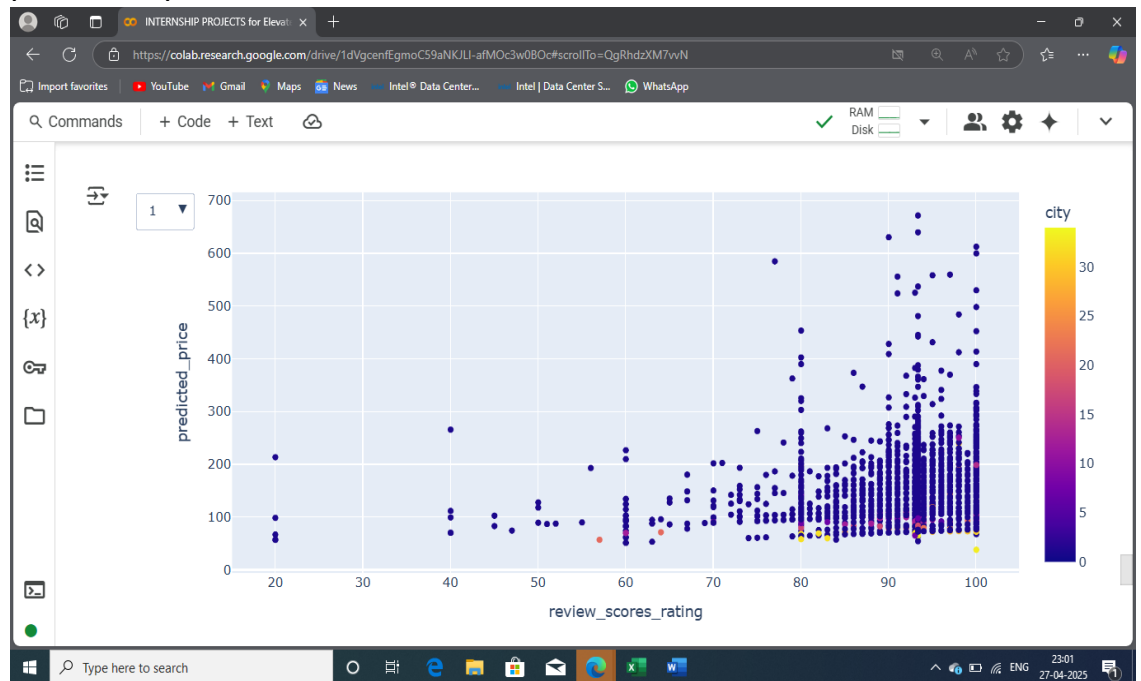
Features.to_csv('airbnb_pricing_for_tableau.csv', index=False)

[23]: # Print the predicted prices
print(Features['predicted_price'])
```

○

5. Visualization:

- Created interactive graphs using Plotly in Colab.
- Developed a Tableau dashboard with filters and sliders for predicted prices.



6. Recommendations: Suggested pricing strategies based on location, listing quality, and seasonal trends.

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

# Example features and target
X = df[['review_scores_rating', 'bedrooms', 'bathrooms']]
y = df['predicted_price']

# Split into train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train model
model = LinearRegression()
model.fit(X_train, y_train)

# Predict
y_pred = model.predict(X_test)

# Model accuracy
from sklearn.metrics import r2_score
r2 = r2_score(y_test, y_pred)
print("R2 Score:", r2)
```

R2 Score: 0.9317940682259636

Conclusion

The project successfully demonstrates how machine learning and visualization tools can be used to recommend dynamic pricing for Airbnb listings. By leveraging review scores, property features, and location data, hosts can be empowered to adjust their pricing more scientifically. The project showcases practical skills in data analysis, predictive modeling, and dashboard development, providing valuable insights for real-world Airbnb operations.