Ride-Sharing Wars: An Analysis of Uber and Lyft Cab Prices

Context

The purpose of this project is to analyze and visualize the cab prices of ride-sharing services Uber and Lyft.

The project aims to provide insights into the pricing strategies of Uber and Lyft by analyzing data on ride prices and other details for both services in various cities in the United States.

What customer can benefit?

- Cost-effective ride selection
- Better ride planning
- Improved transparency
- More efficient ride-sharing

Content

- Data Collection
- Data Cleaning and Preprocessing
- Exploratory Data Analysis
- Statistical Analysis
- Visualisation

Data Collection

Collect data on ride prices and other details for both Uber and Lyft in various cities across the United States.

We used the dataset from Kaggle.

https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices?resource=download

Data Cleaning and Preprocessing

This task involves tasks such as removing duplicates, imputing missing values, and converting data types as needed.

Exploratory Data Analysis

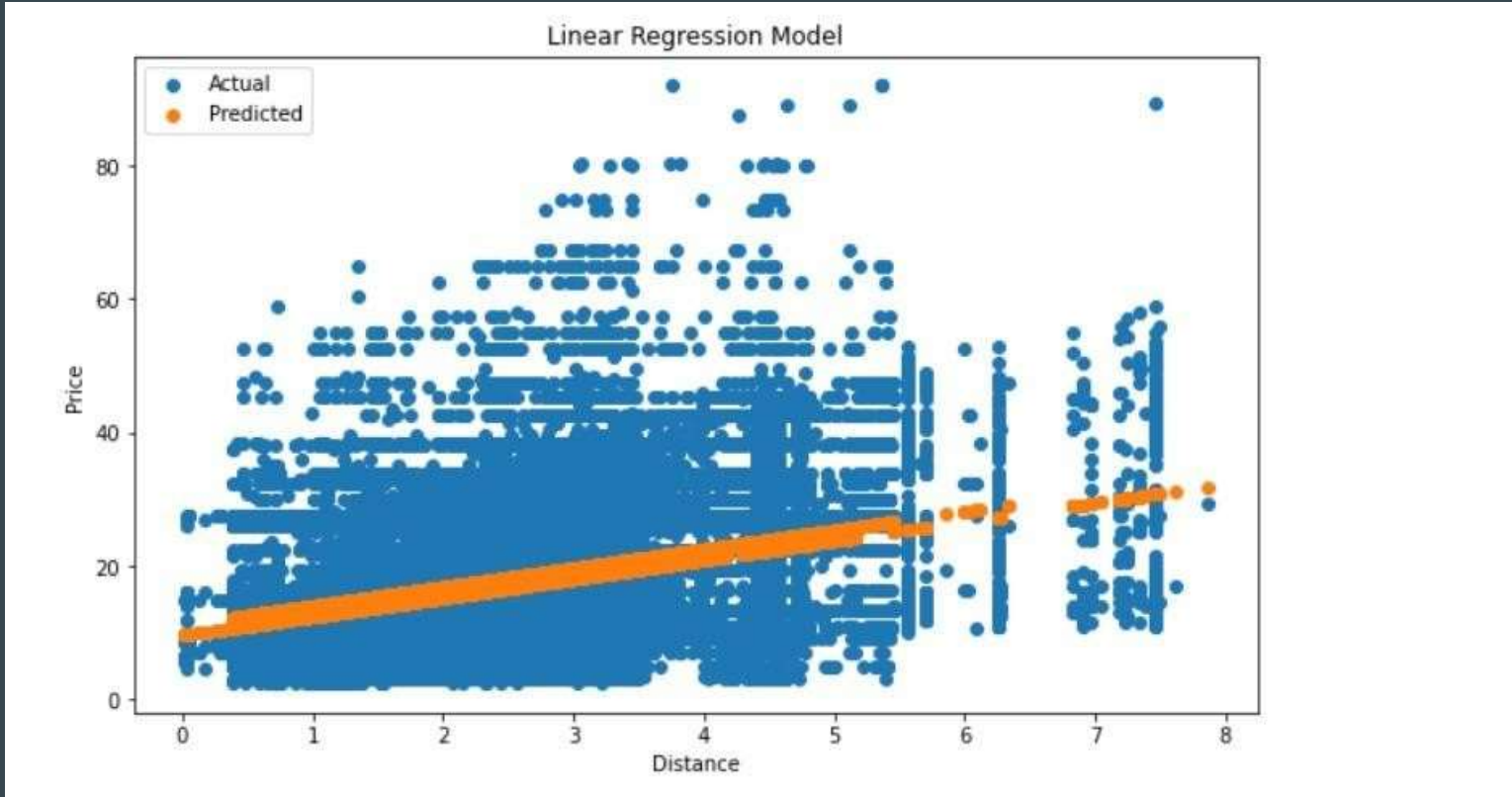
This involved visualizing the data in various ways to identify patterns and trends, such as how ride prices vary by time of day or location.

Statistical Analysis

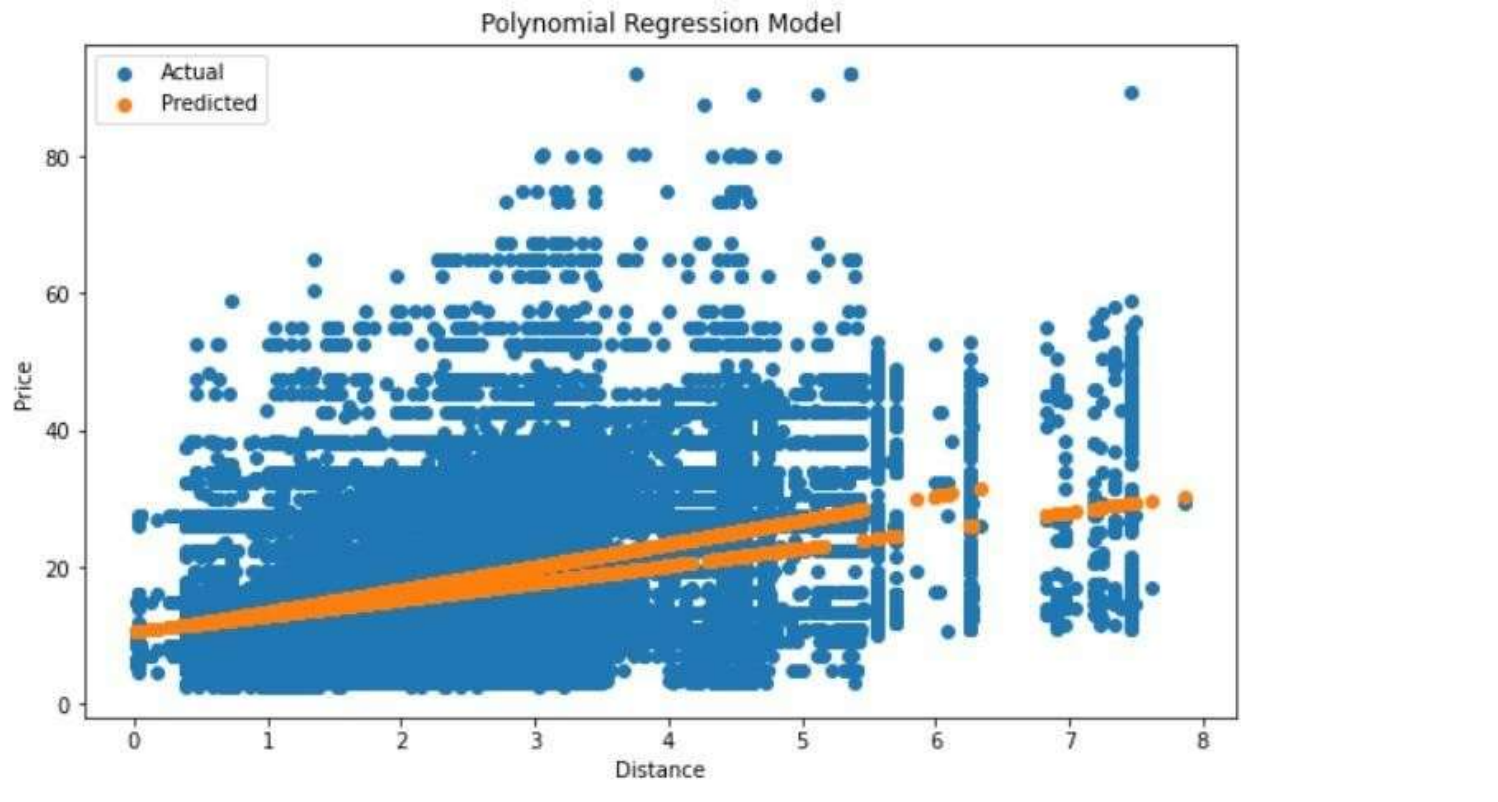
statistical analysis was performed on the data to further investigate the relationships between variables and to test hypotheses

Visualisation

Visualisation



Visualisation



jupyter Uber and lyft Price Prediction Last Checkpoint: 04/14/2023 (autosaved)

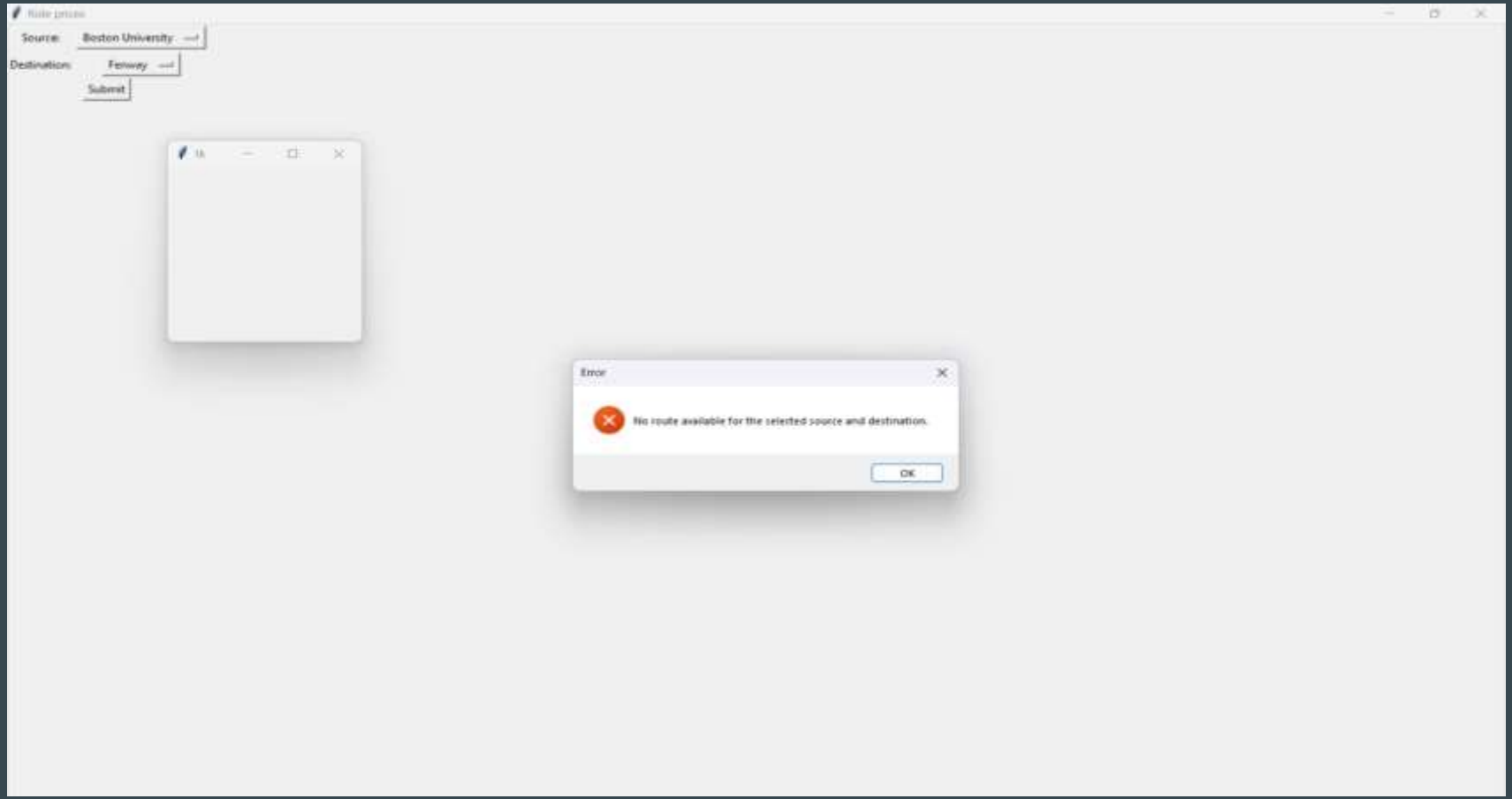
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
# Load the dataset
df = pd.read_csv('cab_rides.csv')

# Get unique sources and destinations from the dataset
sources = df['source'].unique()
destinations = df['destination'].unique()

def get_prices():
    # Get the selected source and destination from the dropdown menus
    source = source_var.get()
    destination = destination_var.get()

    # Filter the dataset to include only rides with the given source and destination
    filtered_df = df[(df['source'] == source) & (df['destination'] == destination)]

    # Check if the filtered dataframe is empty or not
    if filtered_df.empty:
        # Display an error message
        tk.messagebox.showerror(title='Error', message='No route available for the selected source and destination.')
        return

    # Calculate the average price for each cab type
    avg_prices = filtered_df.groupby('cab_type')['price'].mean()

    # Determine the best cab type and its average price
    best_cab_type = avg_prices.idxmin()
    best_price = avg_prices.min()

    # Get the name of the best vehicle
    best_vehicle = filtered_df[filtered_df['cab_type'] == best_cab_type]['name'].iloc[0]

    # Display the results in a message box
    result = "Source: {}\nDestination: {}\n\nLyft price (Lyft): ${:.2f}\nUber price (Uber): ${:.2f}\n\nBest price: ${:.2f} ({}, {
        source, destination, avg_prices.get('Lyft', 0), avg_prices.get('Uber', 0), best_price, best_vehicle, best_cab_type)
    tk.messagebox.showinfo(title='Results', message=result)
```


Source:
Boston University
Destination:
Fenway
Submit
Error
No route available for the selected source and destination.
OK

The feature "no route available" helps when the app cannot find a feasible route between the source and destination the user selects. It can happen for various reasons, such as road closures, traffic congestion, or a lack of available drivers.

When this feature is triggered, the app will display an error message informing the user that no route is available for the selected source and destination. The user must then choose an alternative source or destination or wait for the situation to resolve before trying again.

Conclusion

- **Pricing strategies -** prices tend to be higher during peak hours and in areas with high demand.
- **Cost-effectiveness -** by providing real-time information on estimated ride prices, an app that incorporates the insights gained from this project could help users choose the most cost-effective option
- **Transparency -** allows customers to make more informed decisions and have greater trust in the services they are using.
- **Efficiency -** an app that incorporates the insights gained from this project could suggest ride-sharing with other passengers going to similar destinations, which would reduce the cost of the ride for everyone involved.

Thankyou