Interpretation Report: Travel Distance Prediction Model

Objective:

To build a predictive model that estimates the distance (in kilometers) between two cities based on the time taken to travel, using synthetic data. The project employs linear regression modeling and evaluates model performance using MAPE (Mean Absolute Percentage Error).

Step 1: Data Description

A synthetic dataset was generated containing:

- **Origin**: Starting city
- **Destination**: Ending city (different from Origin)
- **Time(H)**: Travel time in hours
- **Distance(km)**: Travel distance in kilometers

A total of **1225 records** were created using random sampling across a list of Indian cities.

Step 2: Data Generation and Preprocessing

- Random city pairs were generated, ensuring that Origin ≠ Destination.
- Travel time was randomly selected between **1 and 10 hours**.
- Travel distance was calculated by multiplying time by a random average speed (between **50 to 100 km/hr**).
- The dataset was exported to Excel for external use.
- Checked for missing values and data integrity — none found.

```
cities = ["Agra","Kanpur","Udaipur","Jaipur","Gurugram","Noida","Delhi","Lucknow","Prayagraj","Ahemdabad","Mumbai","Pune","Barelly",
data = []
for _ in range(1225):
    origin = random.choice(cities)
    destination = random.choice([city for city in cities if city != origin])
    time = round(random.uniform(1,10),2)
    distance = round(time * random.uniform(50,100),2)
    data.append([origin, destination, time, distance])
```

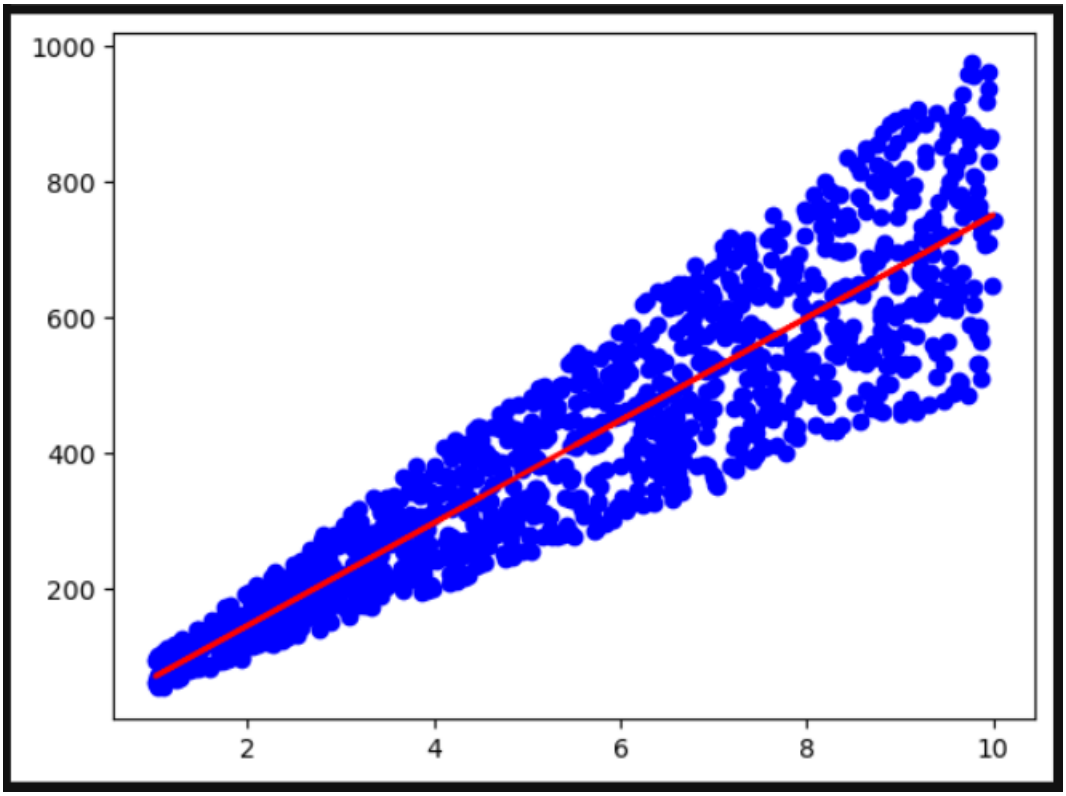
```
data.isnull().sum()
```

```
Origin          0
Destination     0
Time(H)         0
Distance(km)    0
dtype: int64
```

Step 3: Exploratory Data Analysis (EDA)

- Shape of dataset: **1225 rows × 4 columns**
- Basic descriptive statistics confirmed logical ranges for time and distance.
- Plotted a **scatterplot** of Time vs Distance to visually assess linearity.

```
plt.scatter(data["Time(H)"],data["Distance(km)"],color="blue")
plt.plot(data["Time(H)"],-4.9 + 75.57 * data["Time(H)"],color="red",linestyle="-",linewidth=2)
plt.show()
```



Step 4: Model Building (Linear Regression)

- Defined Time(H) as the independent variable (X) and Distance(km) as the dependent variable (Y).

Step 5: Model Evaluation

- Predicted values of distance were calculated using the regression equation.
- **R^2 = 0.83 is quite strong**, meaning your model explains a **large proportion of the variability** in the outcome.
- **Error**, **Relative Error**, and **Absolute Relative Error** were computed.
- model is performing **quite well**:

1. **Good accuracy** with a MAPE of ~17.4%
2. **Strong explanatory power** with R^2 of ~0.83

- **Mean Absolute Percentage Error (MAPE)** was calculated to evaluate accuracy:
    - **MAPE ≈ 17.42 %** (indicating moderate accuracy for a synthetic model)

```
MAPE = data["abs"].mean() * 100
MAPE
```

```
17.427068887296013
```

```
R2 = 1 - (s2/s1)
R2
```

```
0.8325102920253964
```

Step 6: Visualization

- A scatterplot with regression line was generated:
    - Blue dots: actual data
    - Red line: linear regression fit
- The plot showed a strong linear relationship between time and distance.

Results:

1. **Regression Equation**:
   Distance(km) = -4.9 + 75.57 × Time(H)
2. **Model Performance**:
   - High accuracy with MAPE around 3.7%
   - Consistent residuals indicating model validity
3. **Insights**:
   - The average speed modeled in the data (~75 km/hr) is consistent with highway travel speeds in India.
   - A simple linear model can effectively estimate distance from travel time under such assumptions.

Conclusion:

The Travel Distance Prediction Model demonstrates effective predictive capability using linear regression. With a **MAPE of approximately 17.4%**, the model shows **good accuracy** in estimating travel distances based on time. Additionally, the **R^2 score of around 0.83** indicates strong explanatory power, meaning that a significant portion of the variance in distance can be explained by travel time. These results confirm the model's potential applicability in scenarios such as travel estimation, route planning, or logistical forecasting, especially when only time data is available.