# About Dataset

I found an ideal dataset for this task, which is based on the demand for cab rides at a given time and the availability of drivers at a given time. I'll do the demand and Supply Analysis for this Dataset to get insights from it.

## Import Libraries

```
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
pio.templates.default = "plotly_white"
```

## Read Dataset

```
data = pd.read_csv('rides.csv')

print(data.head())
```

```
   Drivers Active Per Hour  Riders Active Per Hour  Rides Completed
0                       72                     295            202.0
1                       50                      78             43.0
2                       40                     250            181.0
3                       78                     140            124.0
4                       74                     195            108.0
```

## null values or not (For Data Preprocessing)

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

```
Drivers Active Per Hour     0
Riders Active Per Hour      0
Rides Completed            54
dtype: int64
```

- The dataset has 54 null values in the Rides Completed column. I'll skip these rows and move forward:

```
data = data.dropna()

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

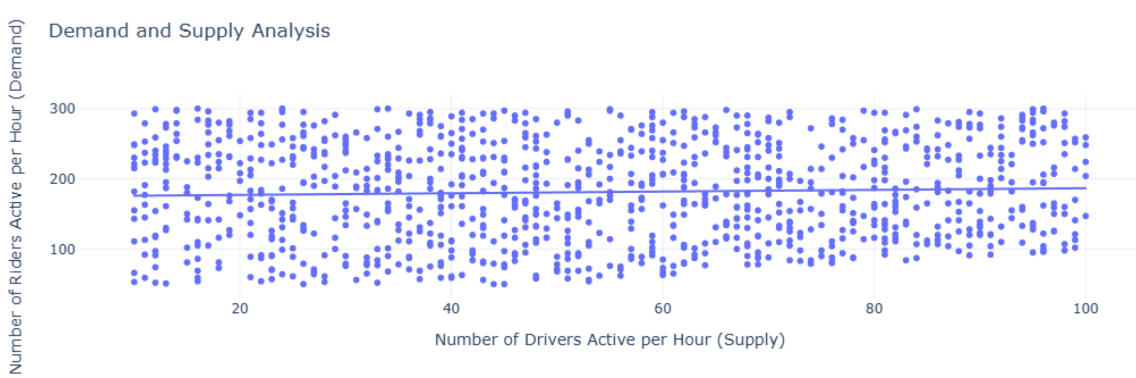
```
Drivers Active Per Hour    0
Riders Active Per Hour     0
Rides Completed            0
dtype: int64
```

- Here two variables need be defined: 1) Rider is a person who take ride means passenger who demand car cab 2) Driver is a person who give ride means driver who supply car cab
- Making variables for demand and supply to find out the relationship.

# Demand and Supply Analysis Graph-

let’s analyze the relationship between the number of drivers active per hour and the number of riders active per hour:

```
demand = data["Riders Active Per Hour"]
supply = data["Drivers Active Per Hour"]

figure = px.scatter(data, x = "Drivers Active Per Hour",
                    y = "Riders Active Per Hour", trendline="ols",
                    title="Demand and Supply Analysis")
figure.update_layout(
    xaxis_title="Number of Drivers Active per Hour (Supply)",
    yaxis_title="Number of Riders Active per Hour (Demand)",
)
figure.show()
```

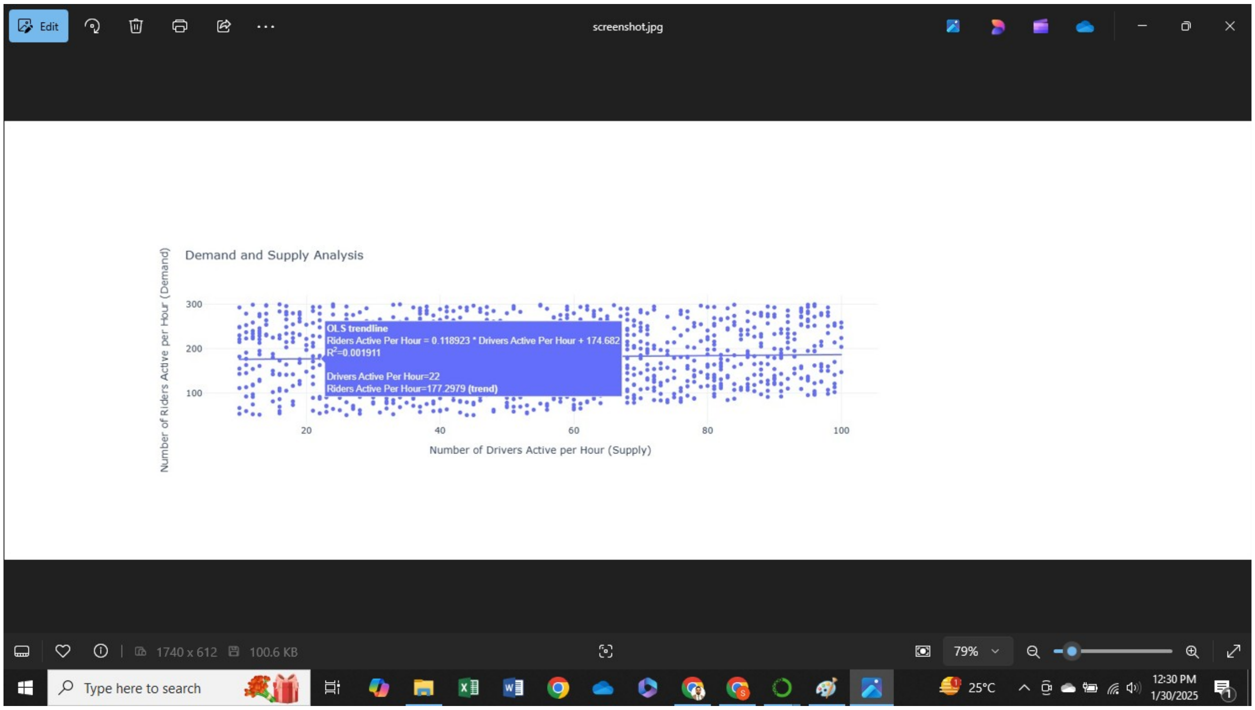



# In the above plot;

- Creates a scatter plot using plotly.express.scatter().
- x (independent variable) is "Drivers Active Per Hour" (supply).
- y (dependent variable) is "Riders Active Per Hour" (demand).
- Adds an Ordinary Least Squares (OLS) regression line (trendline="ols") to see the trend.
- X-axis labeled as supply.
- Y-axis labeled as demand.

# Interpretetion

- Scatter plot shows relationship b/w Active Riders & Active Driver Per Hour
- OLS estimates linear trend b/w Demand (Riders) & Supply (Drivers)




# Insights;

Regression Equation The equation displayed in the tooltip is:

# Riders Active Per Hour

0.118923 × Drivers Active Per Hour

- 174.682 Riders Active Per Hour=0.118923×Drivers Active Per Hour+174.682 where:

0.118923 is the estimated slope ($\beta 1$): This means that for every additional active driver per hour, the number of active riders increases by 0.1189 on average. 174.682 is the intercept ($\beta 0$): When there are zero active drivers, the predicted number of active riders is about 174.68.

# Interpretation of $R^2$ (Coefficient of Determination)

The $R^2$ value = 0.001911, which is very close to zero. This suggests:

The model explains only 0.19% of the variation in the number of riders per hour. The relationship between supply (drivers) and demand (riders) is extremely weak. There might be other significant factors influencing the number of active riders that are not included in this model (e.g., pricing, weather, peak hours, etc.).

# Insights:

Weak Relationship: Since $R^2$ is very low, supply (drivers) is not strongly correlated with demand (riders) in this dataset.

Possible Reasons for Low $R^2$:

- External Factors: Demand might depend more on variables such as ride prices, location, time of day, or seasonal trends.
- Data Noise: There might be high randomness in the data, reducing the explanatory power of supply (drivers).
- Non-linearity: The relationship might not be linear, requiring different regression models (e.g., quadratic, logarithmic).
- Measurement Errors: If data is not accurate or contains outliers, OLS may not capture the true relationship.

# Calculating Elasticity:

The code calculates the elasticity of demand with respect to the number of active drivers per hour, which measures how responsive the number of active riders (demand) is to changes in the number of active drivers (supply).

```
# Calculate elasticity
avg_demand = data['Riders Active Per Hour'].mean()
avg_supply = data['Drivers Active Per Hour'].mean()
pct_change_demand = (max(data['Riders Active Per Hour']) -
min(data['Riders Active Per Hour'])) / avg_demand * 100
pct_change_supply = (max(data['Drivers Active Per Hour']) -
min(data['Drivers Active Per Hour'])) / avg_supply * 100
elasticity = pct_change_demand / pct_change_supply

print("Elasticity of demand with respect to the number of active
drivers per hour: {:.2f}".format(elasticity))
```

```
Elasticity of demand with respect to the number of active drivers per
hour: 0.82
```

# Economic Interpretation

- Elasticity = 0.82 means that demand (riders) is relatively inelastic with respect to supply (drivers).
- Inelastic Demand (0 < E < 1): A 1% increase in the number of active drivers per hour increases demand (riders) by only 0.82%.
- This suggests that increasing the number of drivers has a smaller proportional effect on demand, possibly because:
- Demand is driven more by external factors like price, location, or time of day.
- Riders are not as sensitive to supply changes.

# Calculating Supply Ratio

1. It helps us understand how efficiently available drivers are serving riders.
2. Detect Market Imbalance (Supply vs. Demand) If the supply ratio varies significantly, it could indicate mismatches between rider demand and driver availability. Example: If the ratio drops below 1 in certain hours, there might be too many drivers but not enough riders, leading to idle time.

```
# Calculate the supply ratio for each level of driver activity
data['Supply Ratio'] = data['Rides Completed'] / data['Drivers Active Per Hour']
print(data.head())
```

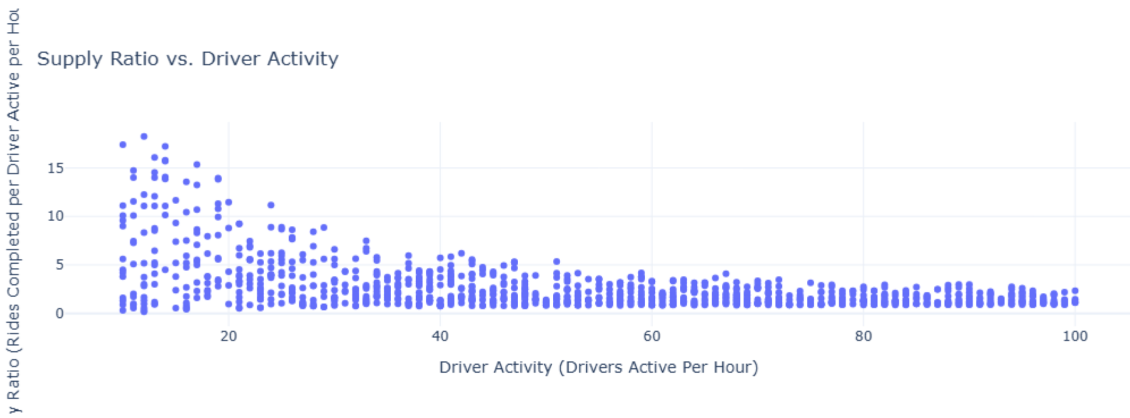
```
   Drivers Active Per Hour  Riders Active Per Hour  Rides Completed  \
0                       72                     295            202.0
1                       50                      78             43.0
2                       40                     250            181.0
3                       78                     140            124.0
4                       74                     195            108.0

   Supply Ratio
0      2.805556
1      0.860000
2      4.525000
3      1.589744
4      1.459459
```

# Interpretation - Measuring Driver Utilization (Efficiency)

A high supply ratio (e.g., 4.5) means that each driver is completing multiple rides per hour, indicating high demand relative to supply. A low supply ratio (e.g., 0.8) suggests that many drivers are idle or there is low demand relative to supply.

# Visualize Supply Ratio

```
fig = go.Figure()
fig.add_trace(go.Scatter(x=data['Drivers Active Per Hour'],
                         y=data['Supply Ratio'], mode='markers'))
fig.update_layout(
    title='Supply Ratio vs. Driver Activity',
    xaxis_title='Driver Activity (Drivers Active Per Hour)',
    yaxis_title='Supply Ratio (Rides Completed per Driver Active per
Hour)'
)
fig.show()
```




# Interpretation of the Scatter Plot: Supply Ratio vs. Driver Activity

This scatter plot visualizes the relationship between the number of drivers active per hour (supply) and the supply ratio (rides completed per active driver per hour).

# Key Observations

Negative Relationship Between Supply and Supply Ratio:

- When fewer drivers are active (< 20 drivers per hour), the supply ratio is high (some points reach 10–15 rides per driver). As more drivers become active, the supply ratio decreases and stabilizes at a lower level. This suggests diminishing returns—as more drivers enter the market, the average number of rides per driver falls. High Variability at Low Driver Activity:
- When the number of active drivers is low, the supply ratio varies widely (ranging from close to 0 to over 15 rides per driver). This might indicate: Demand spikes at certain times, leading to more rides per driver. Random fluctuations due to small

sample sizes. Higher efficiency when there are few drivers competing for available riders. Stabilization at Higher Driver Activity:

- When more than 50 drivers are active, the supply ratio stabilizes at a lower level (~2-5 rides per driver). This suggests that in high-supply situations, the number of rides per driver levels off, likely due to supply exceeding demand.

# Economic Interpretation

Diminishing Marginal Returns to Supply:

- When few drivers are available, each driver completes more rides. When many drivers are available, they compete for the same riders, reducing the number of rides per driver. This follows the classic economic principle of diminishing returns.

Market Equilibrium Insights:

- If the supply ratio is too low, there may be oversupply—drivers might earn less per hour. If the supply ratio is too high, there may be excess demand—longer rider wait times.

Potential Pricing Implications:

- Surge pricing could be beneficial when supply is low to encourage more drivers to enter. Driver incentives may not be necessary when supply is already high, as adding more drivers reduces individual driver earnings.