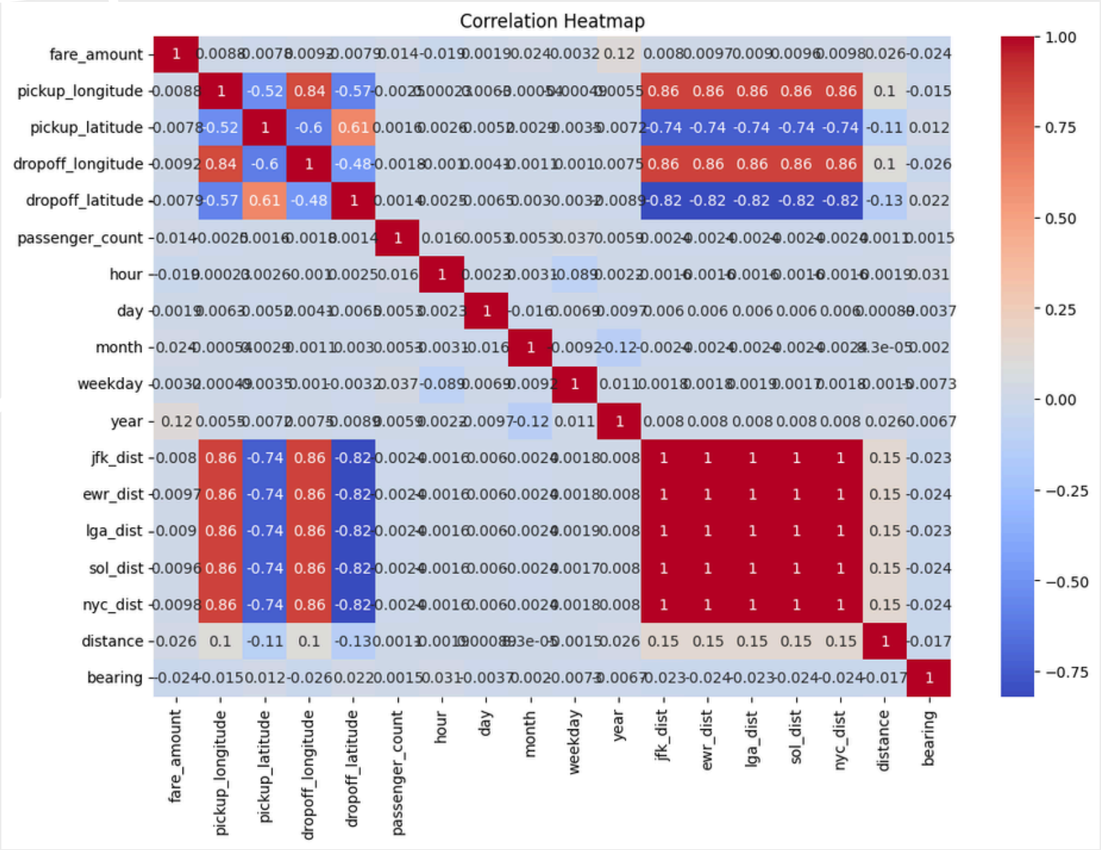
Title: Exploratory Data Analysis of NYC Taxi Fare Dataset

OBJECTIVE

The main objective of this project is to explore a dataset of NYC taxi rides
to discover insights related to:
Fare amounts
Trip times and hours
Weather impact
Passenger behavior patterns

CORRELATION HEATMAP

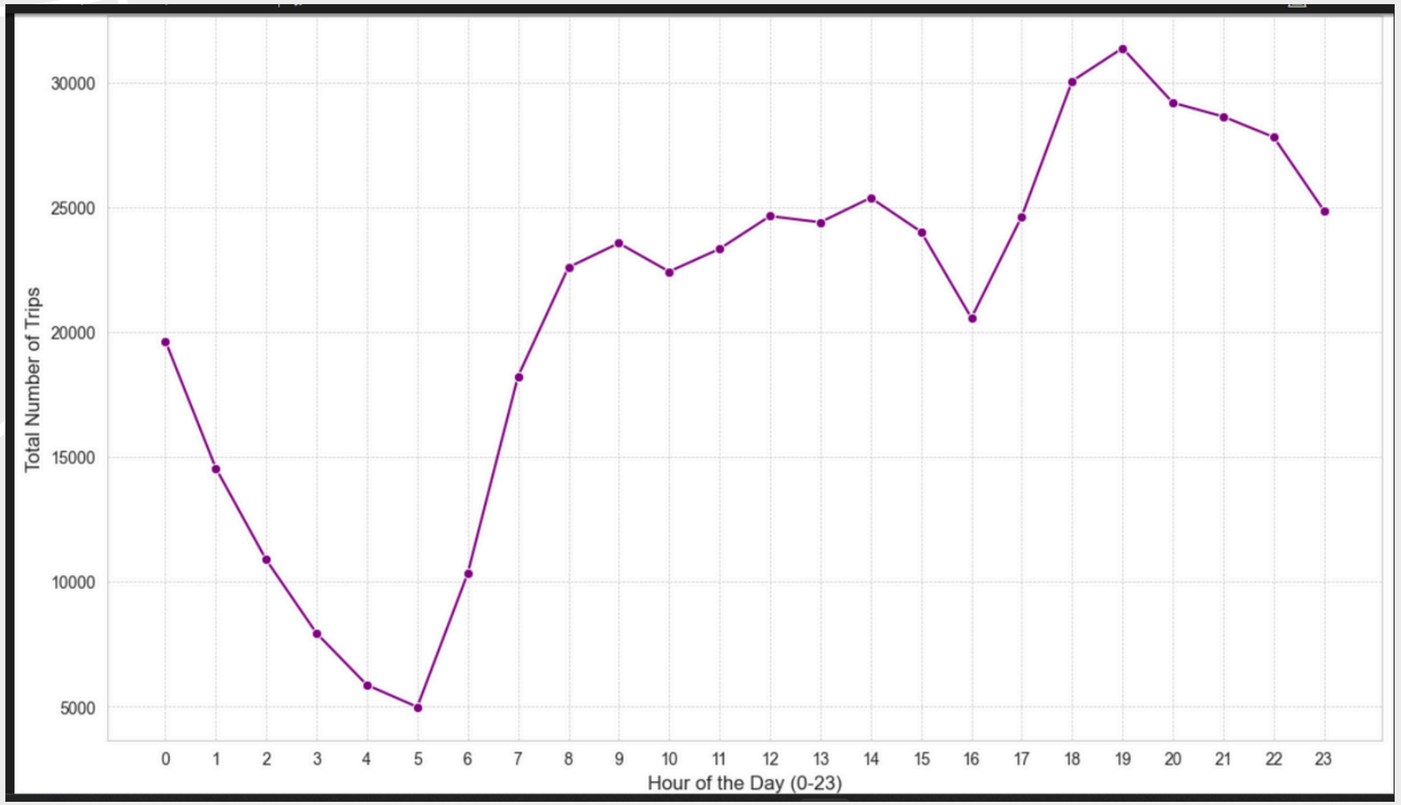


This heatmap shows the correlation between numerical features in the dataset

Strong positive correlation appears in red, and strong negative correlation appears in blue.

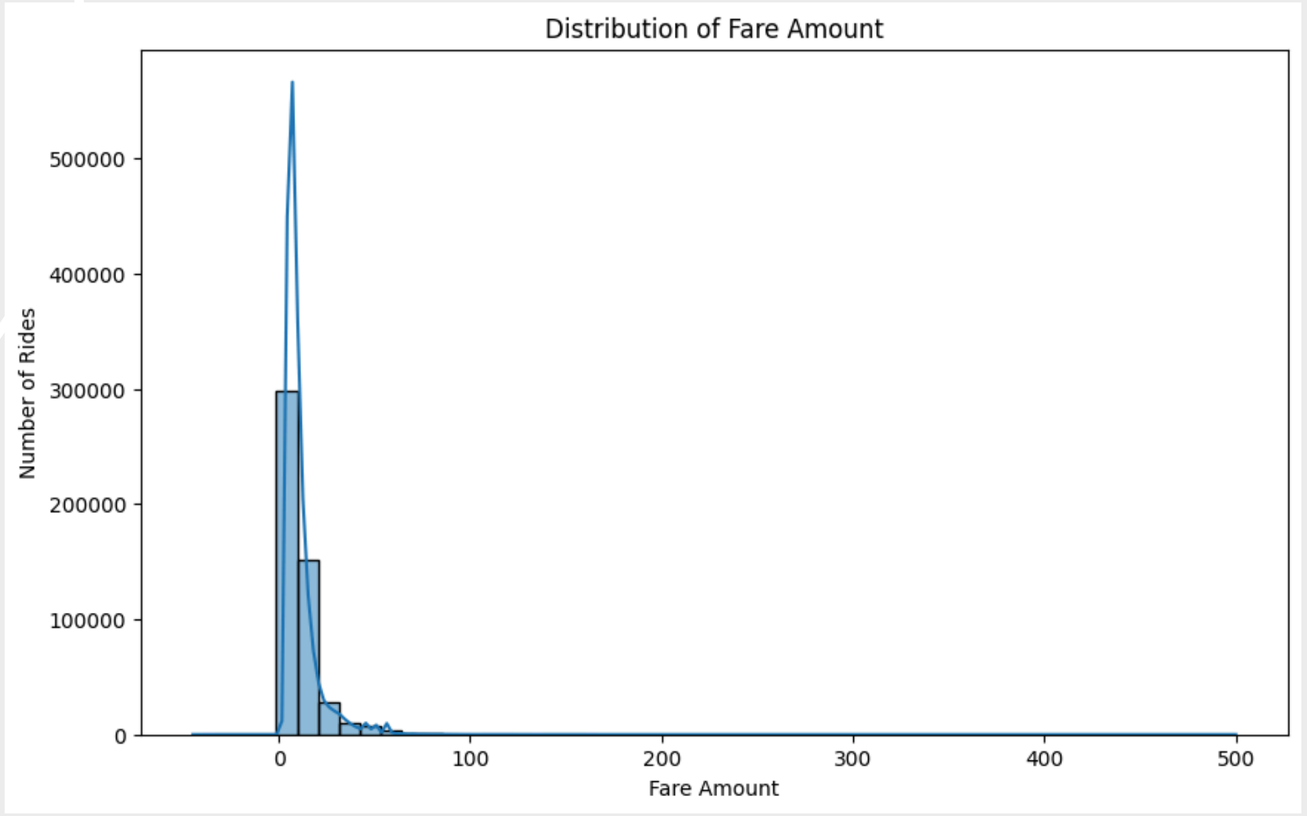
It helps identify which variables are strongly related, such as total_amount and fare_amount

TRIPS DISTRIBUTION BY HOUR



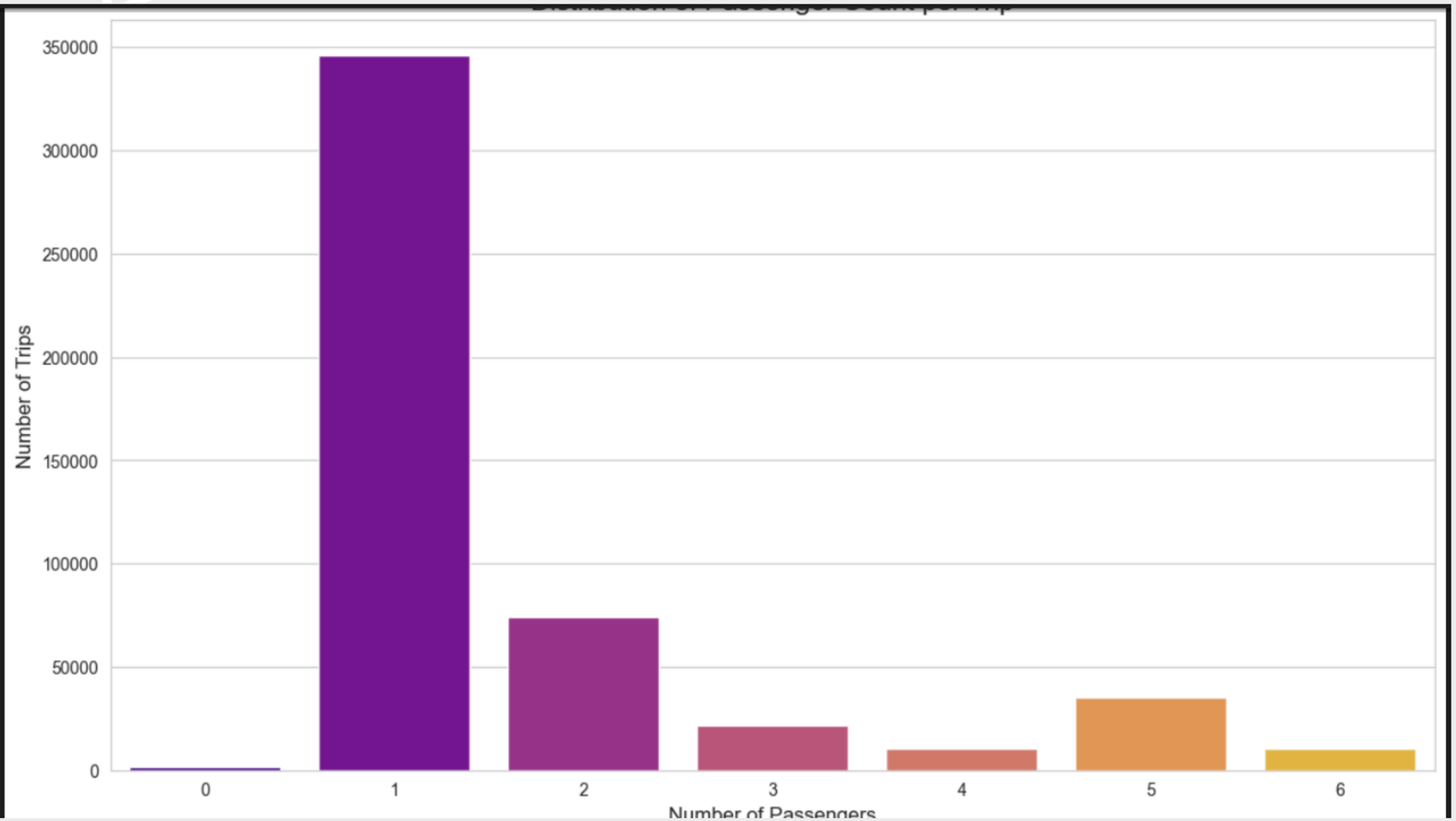
this line plot shows the total number of taxi trips for each hour of the day from 0 to 23 we can observe clear peaks during rush hours especially in the morning and evening this insight helps understand demand patterns throughout the day

FARE AMOUNT DISTRIBUTION



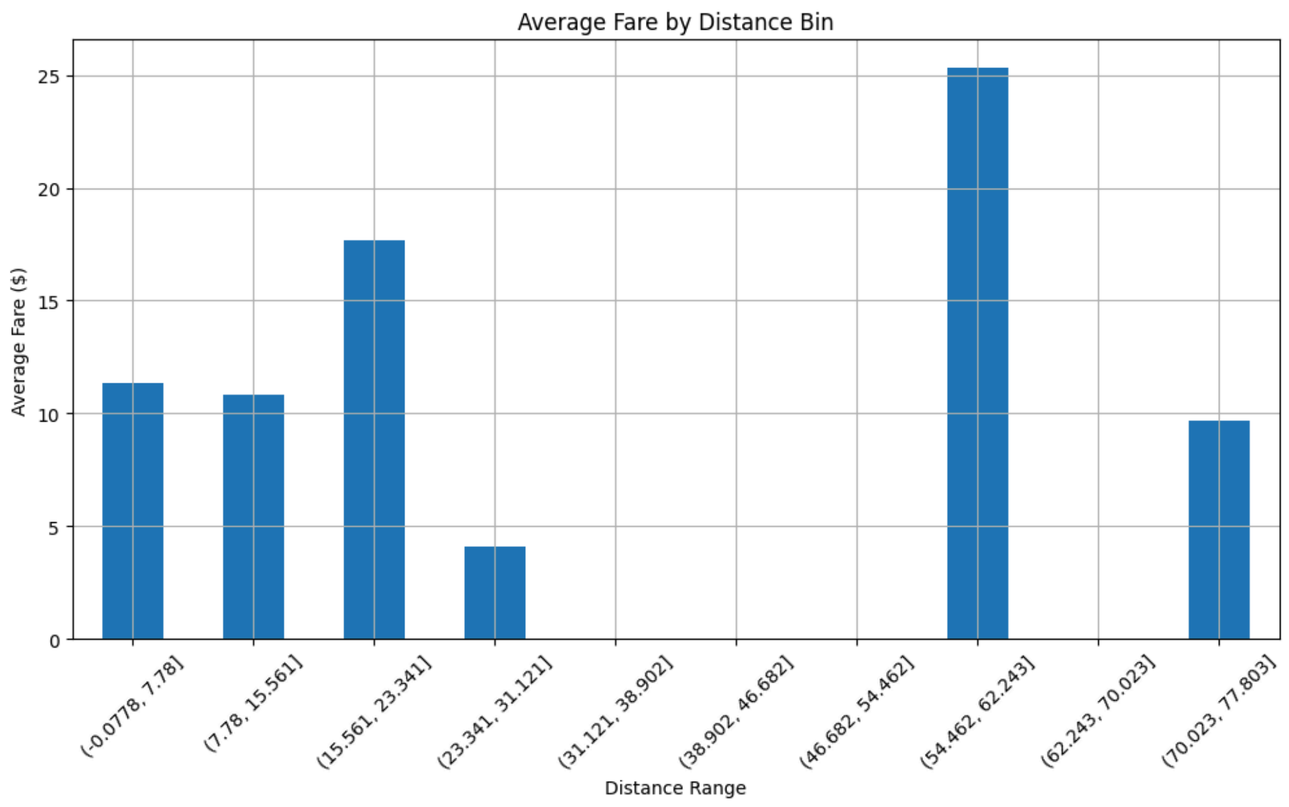
this histogram shows the distribution of fare amounts for all rides most fares fall within a specific range which gives insights into typical pricing patterns outliers on the higher end may indicate long distance trips or additional charges

CATEGORICAL COLUMNS



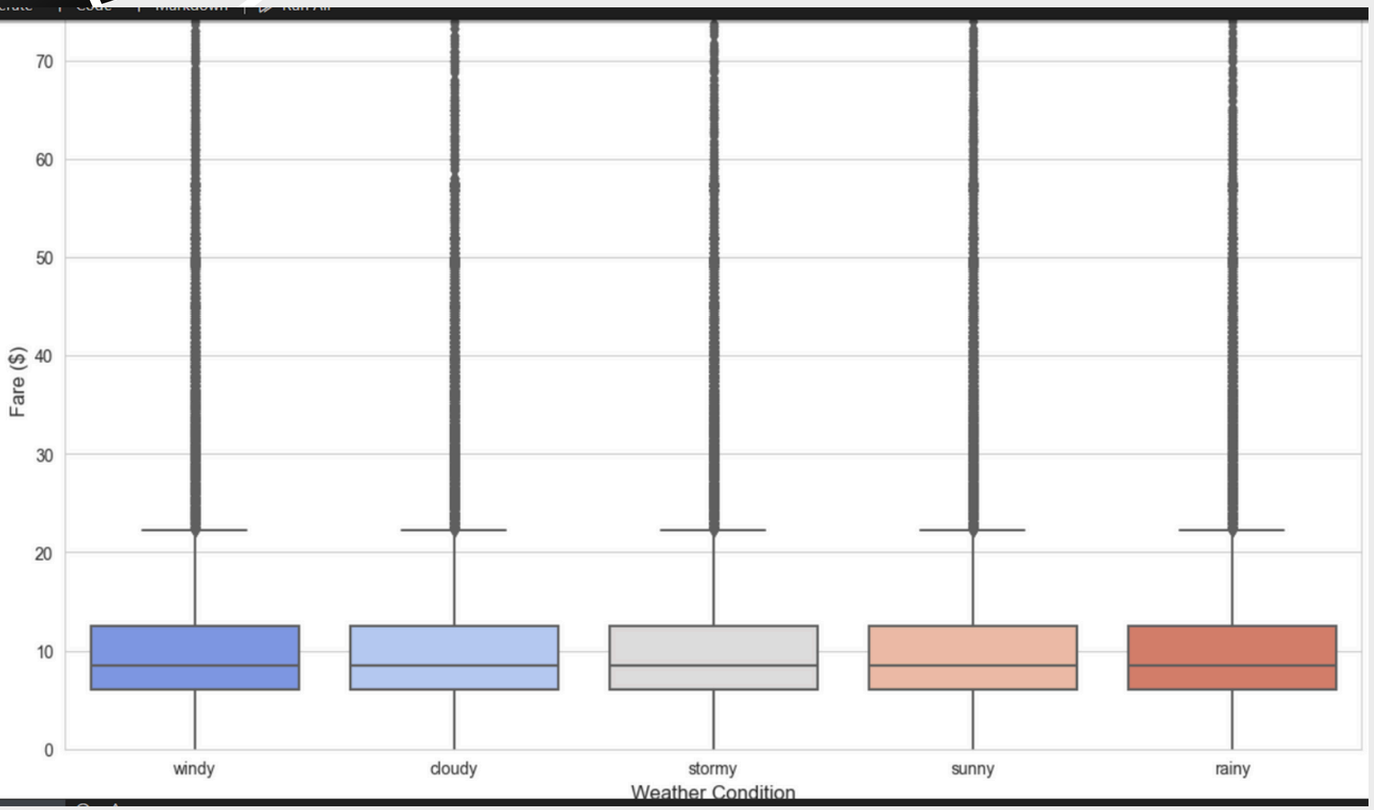
this bar chart shows the distribution of the number of passengers in each taxi trip most trips have one passenger which suggests that people usually travel alone trips with more than three passengers are less common and may indicate group rides or shared taxis

AVERAGE FARE BY DISTANCE BIN



this bar chart shows the average fare based on the distance of each trip the distance was calculated using the pickup and dropoff coordinates then divided into ten equal bins we can see that as the distance increases the average fare also increases which is expected since longer trips cost more

fare amount distribution by weather condition



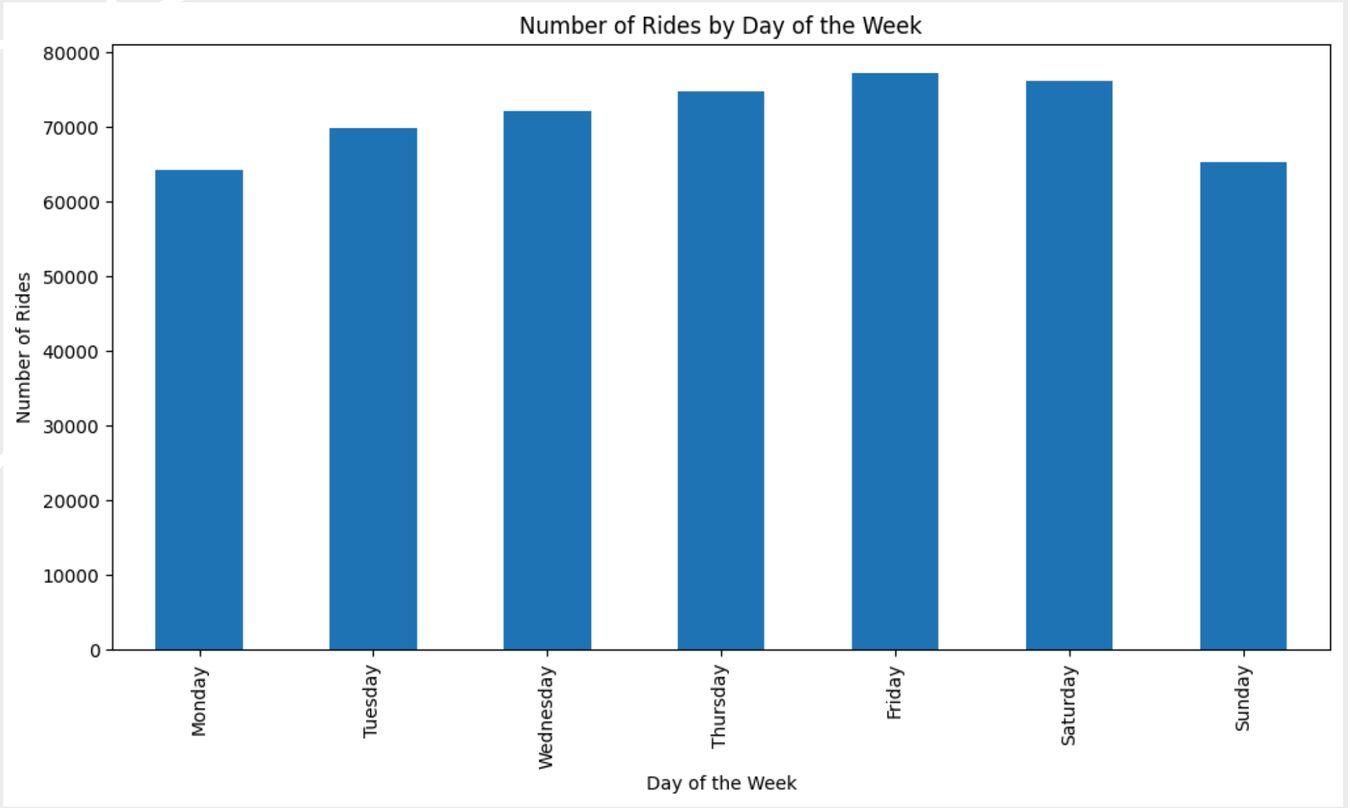
this box plot shows the variation in fare amount across different weather conditions

it highlights how weather can influence the fare — for example, fares may rise during bad weather

the box plot gives us a clear view of the range, median, and outliers of fares in each condition

extreme weather may lead to higher variability in prices

number of rides by day of the week



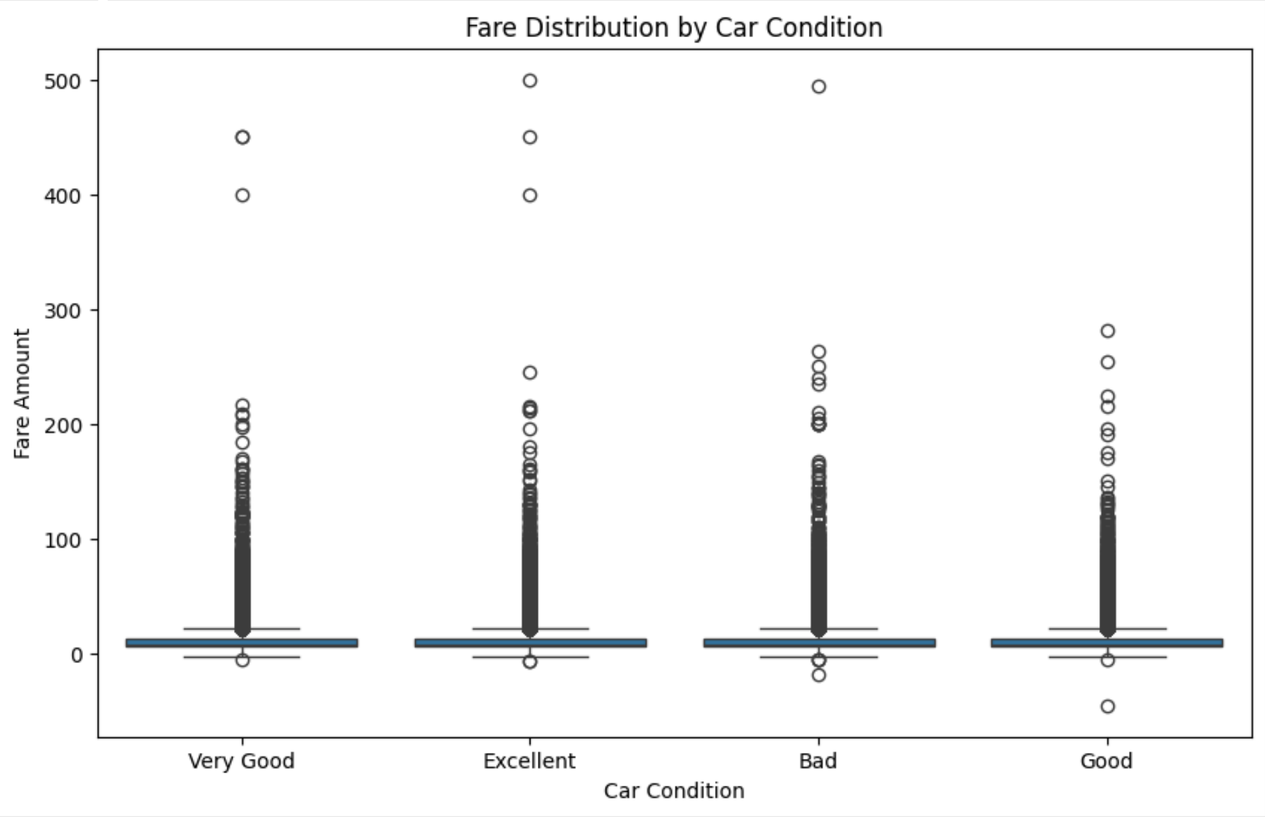
this bar chart shows the number of taxi rides for each day of the week

it helps identify the busiest and least busy days

for example, high ride counts on Friday and Saturday may indicate weekend demand

this information is useful for scheduling drivers or planning promotions

FARE DISTRIBUTION BY CAR CONDITION



this box plot visualizes how fare amounts vary with car condition

we can observe whether better car conditions lead to higher or more consistent fares

the box shows the middle 50% of the data, while the lines (whiskers) show the overall spread

outliers indicate unusually high or low fares

THANK YOU

www.reallygreatsite.com