



PROJECT I

UBER V/S WEATHER

NEW YORK CITY, 2009-14

**SOURCED THE
DATA OF 200K
ROWS OF
UBER
BOOKING IN
NEW YORK
CITY FROM
JANUARY
2009 – JUNE
2015**



**WE FOUND INCORRECT
VALUES IN THE DATA**

VALIDATED THE DATA:
REMOVED THE OUTLIERS
USING: IQR RULE FOR
LATITUDES AND LONGITUDES

REMOVED ROWS WITH
NEGATIVE AND NEGLIGIBLE
FARE VALUES

REMOVED ROWS WITH '0'
DISTANCE

DATA FOR 2015 WAS FOR 6
MONTHS; HENCE DECIDED TO WORK
ON THE DATA 2009-2014 FOR
BETTER MONTH-ON-MONTH
COMPARISON, IF NEEDED

RETAINED COLUMNS:

PICKUP DATETIME

PICKUP LATITUDE

PICKUP LONGITUDE

DROPOFF LATITUDE

DROPOFF LONGITUDE

PASSENGER COUNT

FARE AMOUNT

ADDED COLUMNS:

DISTANCE (MILES) USING PICKUP
AND DROP OFF COORDINATES

USING BINS: TIME (INTERVAL),
FARE AMOUNT GROUP, YEAR,
MONTH USING BINS

TIMESTAMP – TO FETCH THE
WEATHER DATA FOR EACH ROW
USING OPENWEATHER API



SAMPLE HISTORICAL API RESPONSE

```
"lat": 52.2297,
"lon": 21.0122,
"timezone": "Europe/Warsaw",
"timezone_offset": 3600,
"data": [
  {
    "dt": 1645888976,
    "sunrise": 1645853361,
    "sunset": 1645891727,
    "temp": 279.13,
    "feels_like": 276.44,
    "pressure": 1029,
    "humidity": 64,
    "dew_point": 272.88,
    "uvi": 0.06,
    "clouds": 0,
    "visibility": 10000,
    "wind_speed": 3.6,
    "wind_deg": 340,
    "weather": [
      {
        "id": 800,
        "main": "Clear",
        "description": "clear sky",
        "icon": "01d"
```

SOURCED THE JSON RESPONSE USING OPEN WEATHER API TO UNDERSTAND THE AVAILABLE PARAMETERS

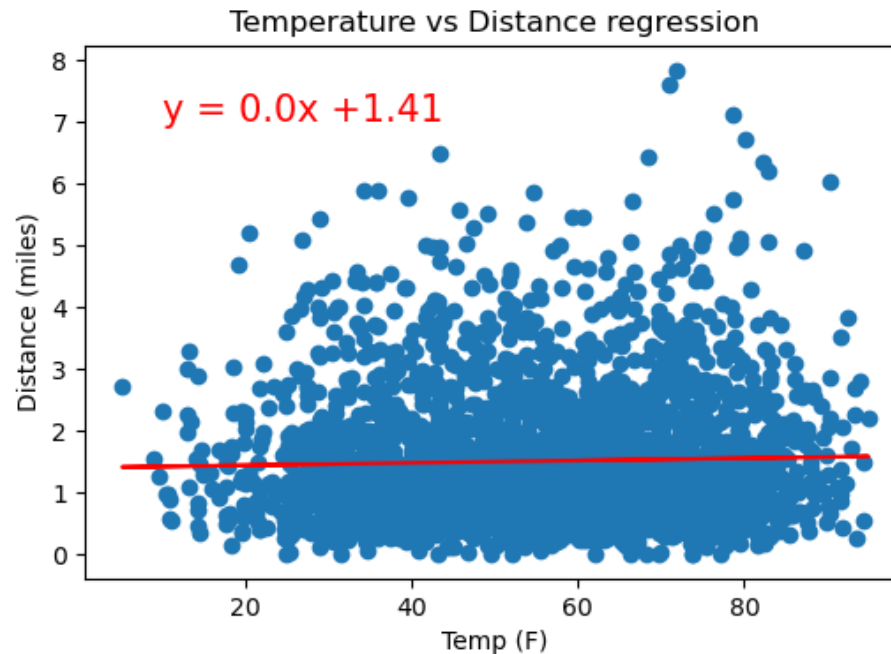
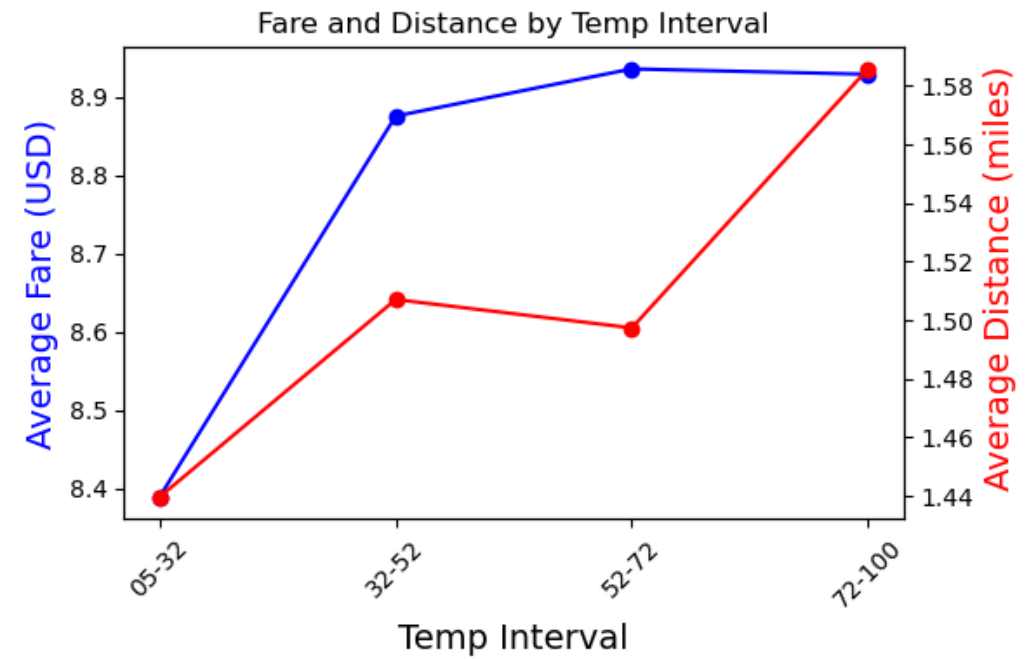
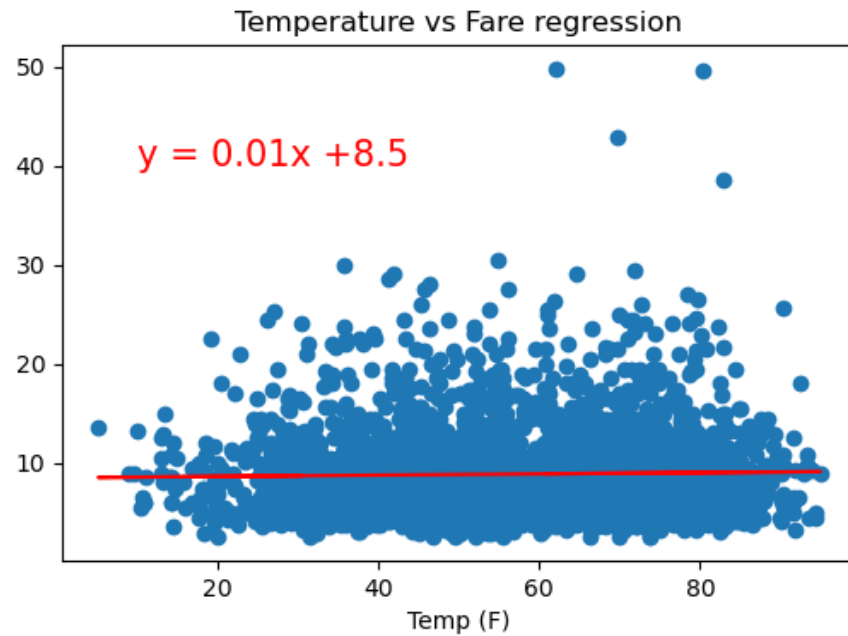
DUE TO API LIMITATION: WE DECIDED TO WORK ON A SAMPLE OF 2900 ROWS WITH ~950 ROWS PER TEAM MEMBER TO WORK ON AND FETCH THE DATA USING TIMESTAMP, PICKUP LATITUDE AND PICKUP LONGITUDE



DATA RETRIEVED FOR EACH ROW FOR:

| | |
|-------------|---------------------|
| CLOUDS | WEATHER DESCRIPTION |
| DEW POINT | WEATHER ICON |
| FEELS LIKE | WEATHER ID |
| HUMIDITY | WEATHER MAIN |
| PRESSURE | WIND DEGREE |
| SUNRISE | WIND SPEED |
| SUNSET | DATA RAIN |
| TEMPERATURE | DATA SNOW |
| VISIBILITY | |

ANALYSIS - EFFECT OF TEMPERATURE ON UBER BOOKINGS

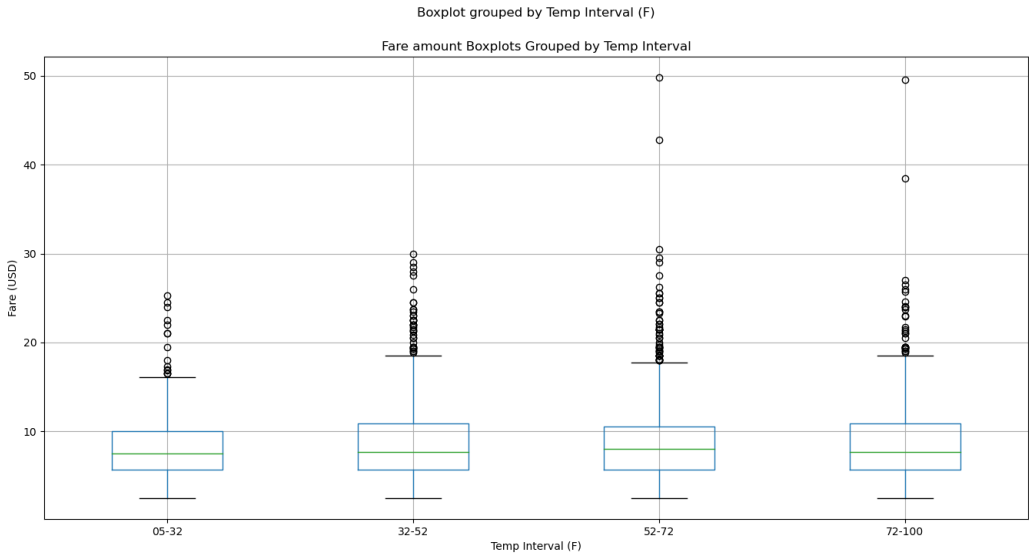
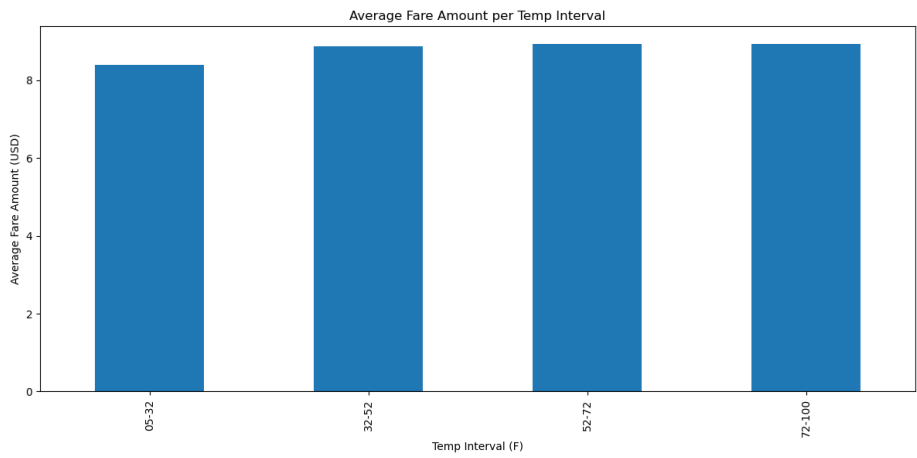


REGRESSION ANALYSIS SHOWS NO / WEAK CORRELATION BETWEEN TEMPERATURE V/S FARE AND DISTANCE

WE DECIDED TO PLOT A DUAL AXIS CHART TO ANALYZE THE TREND BETWEEN FARE AND DISTANCE AS PER THE CREATED TEMPERATURE INTERVALS.

IT SEEMS LIKE THERE IS SOME TREND BETWEEN THE FIRST TWO INTERVALS; HOWEVER, WE WOULD LIKE TO DEEP DIVE AND FIND IF THERE IS SOME CORRELATION WITH THE HELP OF STATISTICAL TESTS.

ANALYSIS – EFFECT OF TEMPERATURE ON UBER BOOKINGS



NULL HYPOTHESIS

THERE IS NO SIGNIFICANT EFFECT OF TEMPERATURE ON THE FARE AND DISTANCE OF UBER

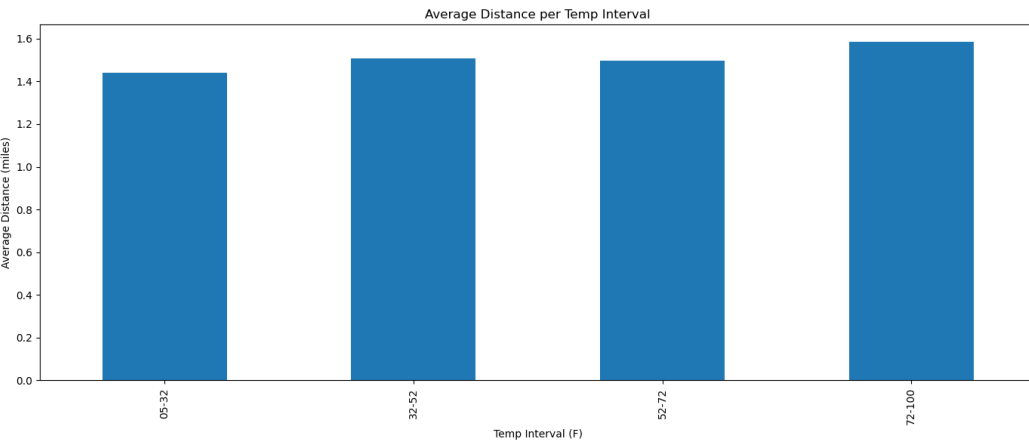
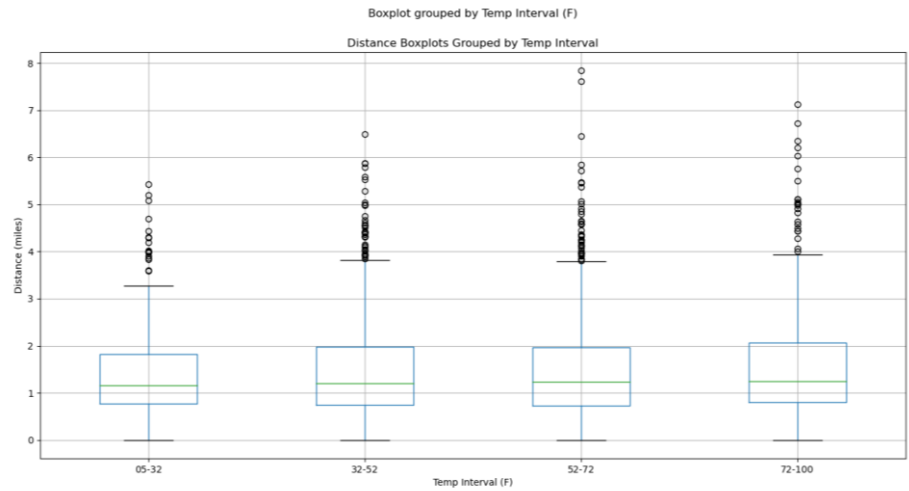
WE PERFORMED ANOVA TEST TO COMPARE THE AVERAGE FARE AND AVERAGE DISTANCE ACROSS TEMPERATURE INTERVALS

P-VALUE FOR AVERAGE FARE =0.313

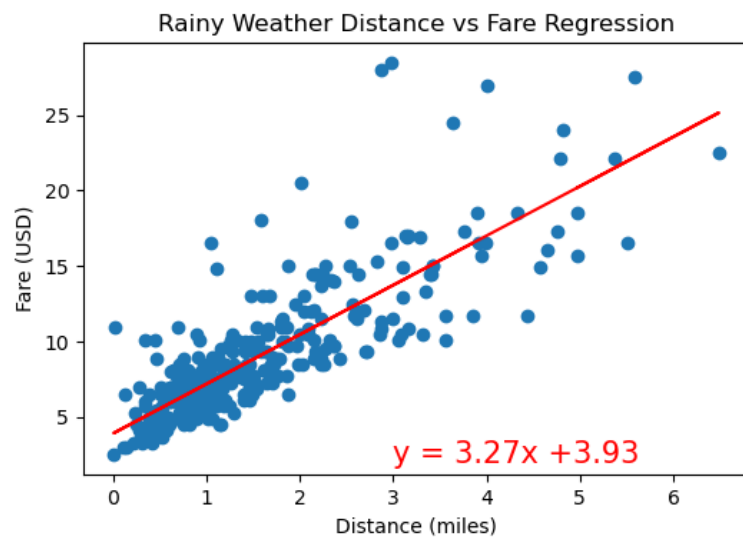
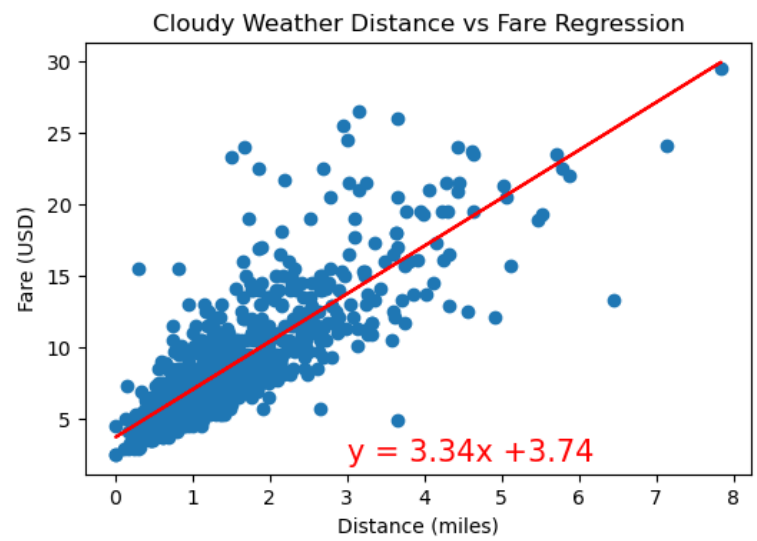
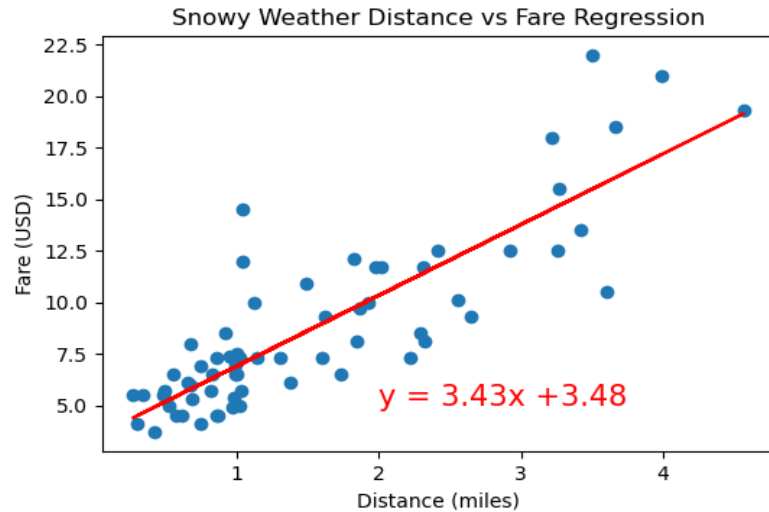
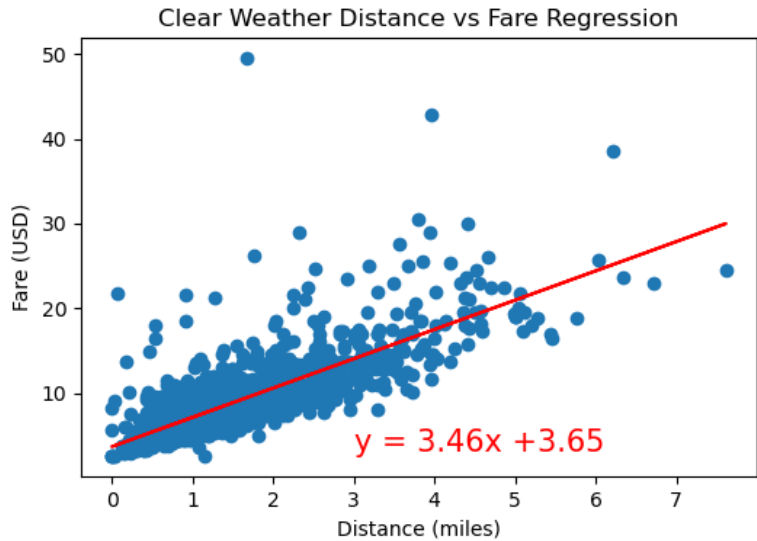
P-VALUE FOR AVERAGE DISTANCE =0.21

WITH THE ANOVA TESTS YIELDING A P-VALUE> 0.05, WE CANNOT REJECT THE NULL HYPOTHESIS.

TEMPERATURE DOES NOT SEEM TO HAVE AN EFFECT ON THE FARE AND DISTANCE OF UBER TRIPS



ANALYSIS - EFFECT OF WEATHER ON UBER BOOKINGS



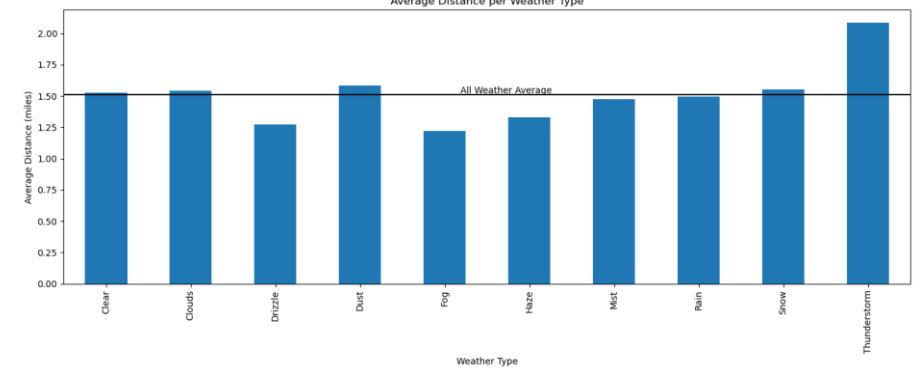
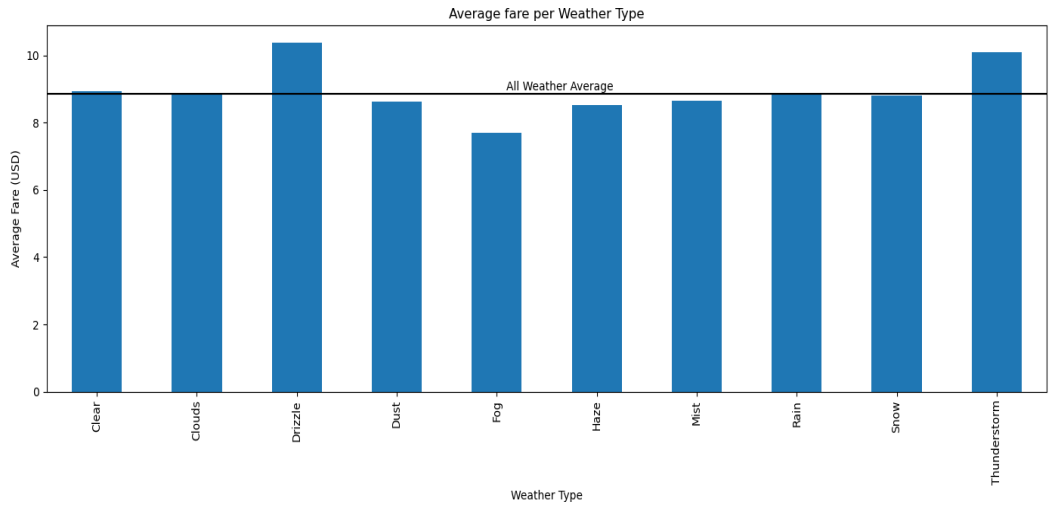
REGRESSION PLOTS SHOW STRONG CORRELATION BETWEEN TRIP DISTANCE AND FARE AS EXPECTED. THE FOLLOWING ARE THE R VALUES:

- CLEAR WEATHER: 0.78
- CLOUDY WEATHER: 0.82
- SNOWY WEATHER: 0.85
- RAINY WEATHER: 0.82

THIS SUGGESTS PRICES MIGHT BE HIGHER DURING CLOUDY, SNOWY AND RAINY WEATHER. FURTHER STATISTICAL ANALYSIS FOLLOWS



ANALYSIS - EFFECT OF WEATHER ON UBER BOOKINGS



NULL HYPOTHESIS: THERE IS NO STATISTICALLY SIGNIFICANT EFFECT OF WEATHER TYPE ON AVERAGE FARE AMOUNT AND AVERAGE TRIP DISTANCE

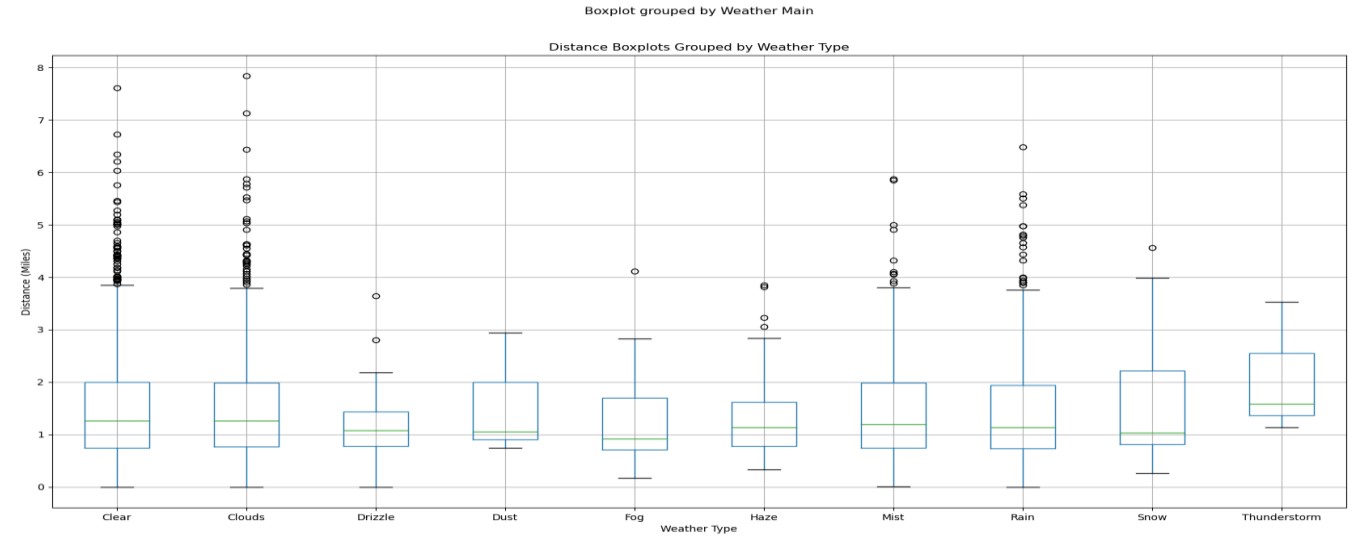
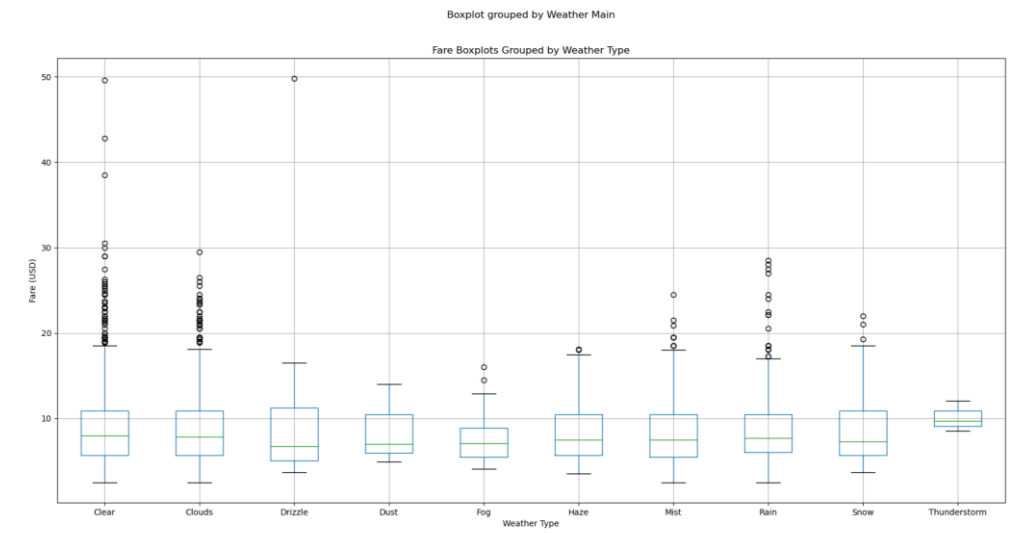
WE CARRIED OUT ANOVA TESTS BETWEEN THE VARIOUS WEATHER TYPES WITH REGARD TO AVERAGE FARE AND AVERAGE DISTANCE

P-VALUE FOR AVERAGE FARE: 0.92

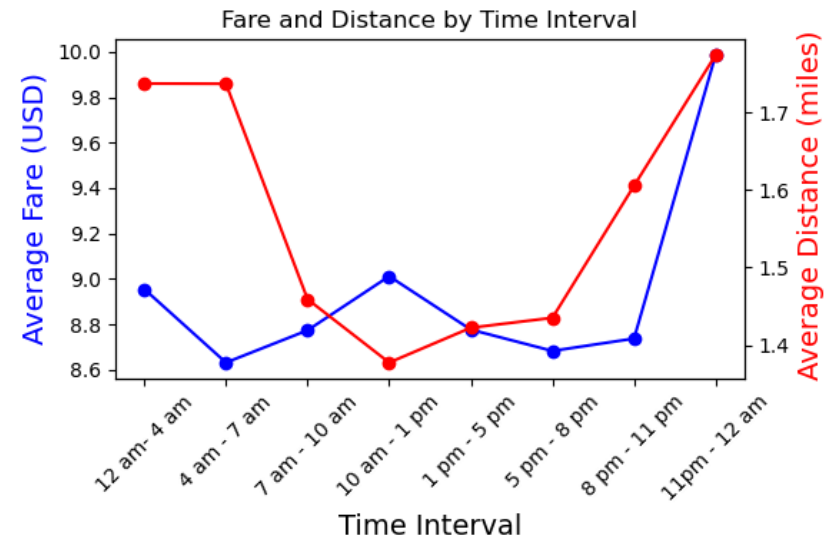
P-VALUE FOR AVERAGE DISTANCE: 0.92

CONCLUSION: THE ANOVA TESTS RETURNED P-VALUES GREATER THAN 0.05 FOR THE DIFFERENT GROUPINGS OF WEATHER TYPES FOR BOTH AVERAGE FARE AND AVERAGE TRIP DISTANCE.

WE COULD NOT REJECT THE NULL HYPOTHESIS/ESTABLISH A FIRM LINK BETWEEN WEATHER TYPE AND FARE AMOUNT OR TRIP DISTANCE.



ANALYSIS - EFFECT OF TIME ON UBER BOOKINGS

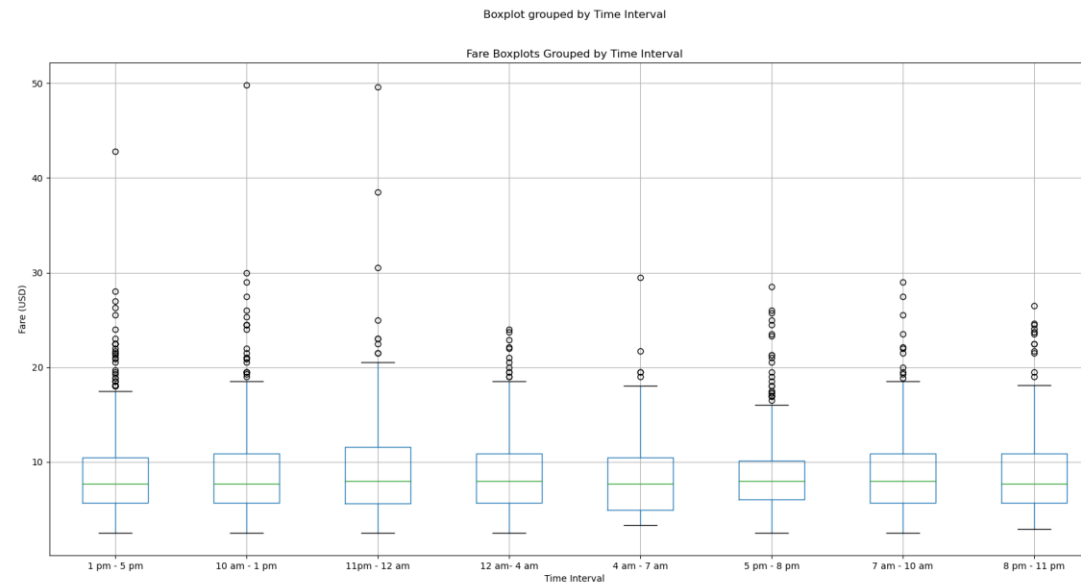
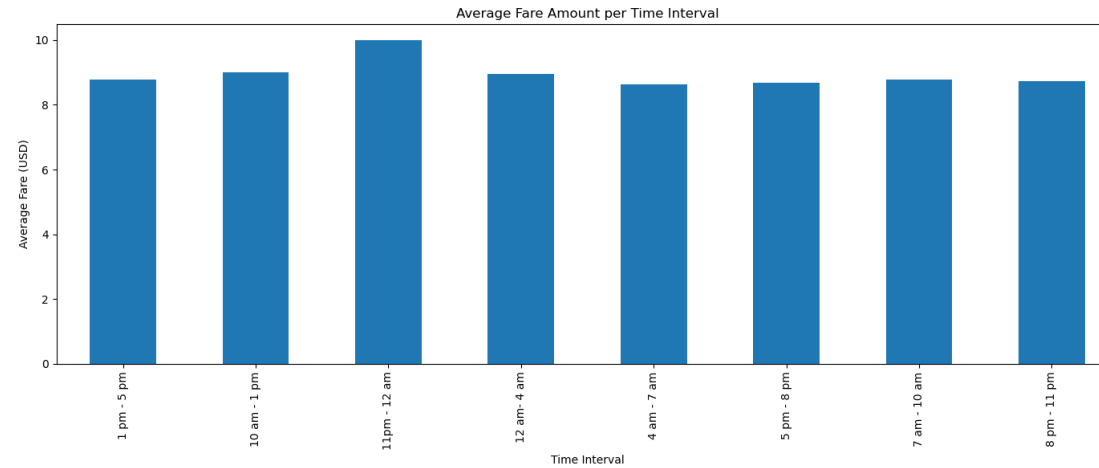


WE DECIDED TO PLOT A DUAL AXIS CHART TO ANALYZE THE TREND BETWEEN FARE AND DISTANCE AS PER THE TIME INTERVALS.

THE GRAPH INDICATES THAT THERE IS A TREND BETWEEN THE TWO VARIABLES.

WE HAVE DONE STATISTICAL ANALYSIS BETWEEN FARE AND DISTANCE INDIVIDUALLY WITH TIME INTERVALS, FOR DIFFERENT WEATHER CONDITIONS TO STUDY IF THERE IS ANY RELATION AMONG THE VARIABLES.

- BAR AND BOX PLOTS FOR FARE AMOUNT PER TIME
- SUMMARY STATISTICS SHOWS THE MEANS AND MEDIANS ARE VERY CLOSE FOR ALL FARE PER TIME INTERVALS



ANALYSIS - EFFECT OF TIME ON UBER BOOKINGS

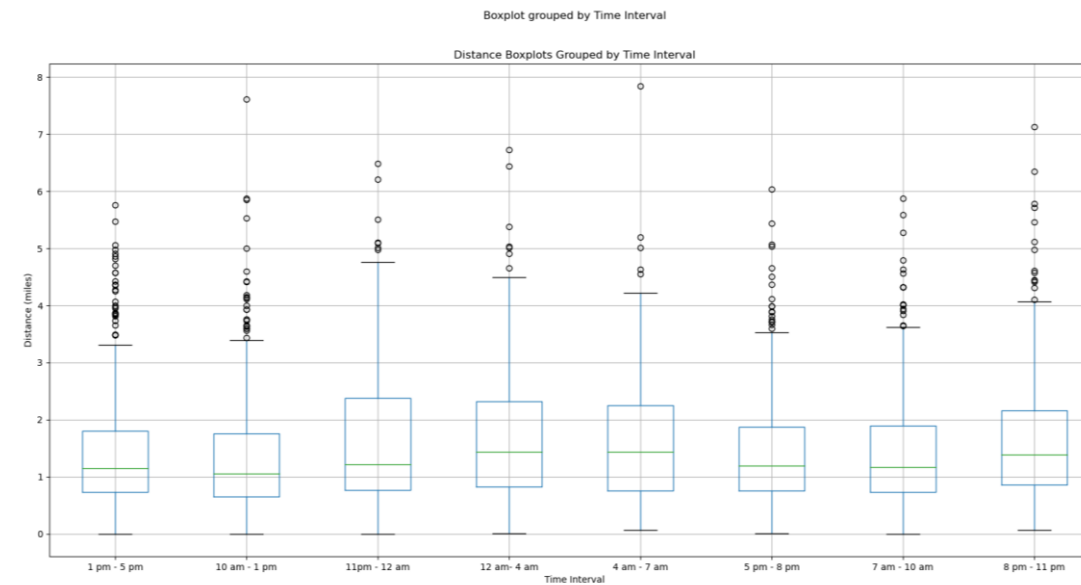
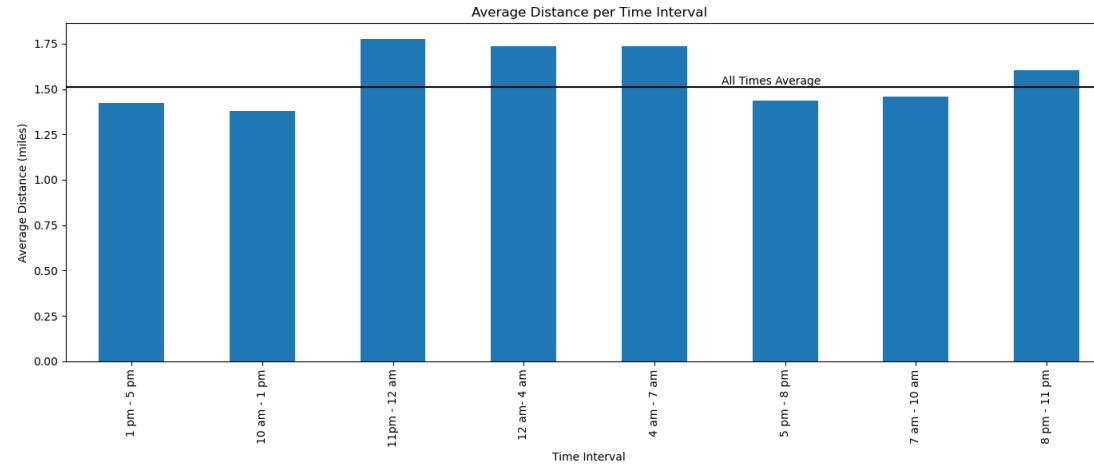
OVERALL CONCLUSION

WE COULD NOT CONCLUDE THAT THERE IS ANY EFFECT OF TEMPERATURE OR WEATHER TYPE ON UBER FARE AMOUNT OR UBER TRIP DISTANCE.

HOWEVER, THERE SEEM TO BE AN EFFECT ON UBER TRIP DISTANCE FOR CERTAIN TIME INTERVALS.

LIMITATIONS

FOR MORE ACCURATE RESULTS, WE SUGGEST TO RUN THIS CODE ON A BIGGER SET OF DATA



NULL HYPOTHESIS: THERE IS NO STATISTICALLY SIGNIFICANT DIFFERENCE IN FARE AMOUNT OR TRIP DISTANCE FOR DIFFERENT TIME INTERVALS.

CONCLUSION FARE AMOUNT: ANOVA TEST P-VALUE: 0.12.SINCE P-VALUE IS GREATER THAN 0.05, WE WERE UNABLE TO REJECT THE NULL HYPOTHESIS.

CONCLUSION TRIP DISTANCE: ANOVA TEST P-VALUE: 0.00000002<0.05, WE REJECT THE THE NULL HYPOTHESIS. FOR FURTHER STUDY, WE CARRIED OUT ONE SAMPLE T-TEST BETWEEN DIFFERENT AVG TRIP DISTANCE PER TIME INTERVALS AGAINST THE POPULATION TRIP DISTANCE. MOST SHOWED STATISTICALLY SIGNIFICANT DIFFERENCE EXCEPT TIME GROUPS 4AM-7AM, 7AM-10AM AND 5PM-8PM

A blurred image of a car's interior, showing the steering wheel and dashboard.

QUESTIONS ARE WELCOME

A solid dark blue rectangular background.

FAHMIDA BILLA

KARAN ANAND

KEVIN MOSWEU

A hand holding a black smartphone. The screen displays the white Uber logo. The background is a blurred car interior.

Uber