Methods of Advanced Engineering

09.01.2024

—

Rajesh Madhipati
Matriculation : 23007086

Introduction

Traveling has become a part of life and most of our day-to-day activities are hugely dependent on traveling, Air travel is a pervasive norm and more than 30 million people use air transport in a single day. Weather is one of the most important features that affect any transport, so we need to know how much impact it causes on a whole.

In this project, for the analysis, we have taken Munich city weather data and people flying from the Munich airport from 2000 to 2023. We try to find the correlation between the number of passengers traveling from Munich airport with respect to the weather.

Methods

Data sources

Munich weather data

Metadata: link

Data: link

Data type: CSV

Description: The monthly weather data set includes the topics of air temperature, humidity, precipitation, and sunshine from 2000 to 2023 in the Munich metropolitan,

Munich Airport air traffic data

Metadata: link

Data: link

Data type: CSV

Transformations

Preprocessing of the Munich weather data

1. Extract the data from the given source and store it in the data folder.
2. Once extracted translate the data from German to English.
3. Drop rows that have null values and mean the important values.
4. Slice the Month column to get the exact month number.
5. Air temperature, Humidity, Precipitation, and Sunshine are arranged row-wise, grouped and split the table row-wise, and then merged with the split data frames.

Preprocessing of the Munich airport air traffic data

1. Extract the data from the given source and store it in the data folder.
2. Once extracted translate the data from German to English.
3. Drop the first 14 rows as they have null values and drop 4 columns that contain information about the previous years and reset the index.
4. Drop the duplicate values insert a new column of the number of passengers and get values from the value column of the data frame.

Combining both datasets

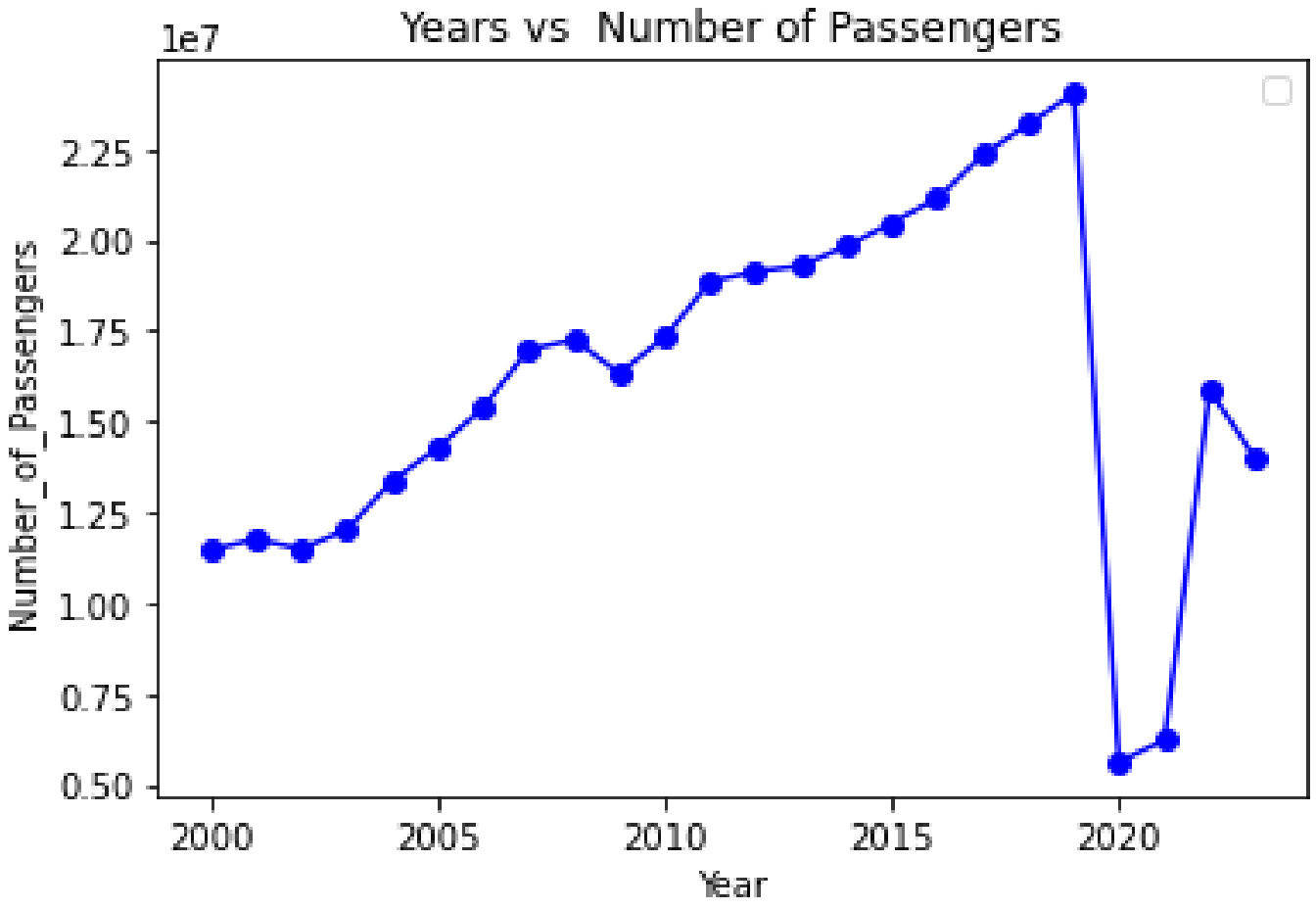
1. Once you are done with preprocessing we can merge both the data sets with Month and year as our primary keys as both these columns are common among these datasets.
2. Once merged remove the overlapping rows and keep only the necessary columns like Year, Month, Humidity, Precipitation, Air Temperature, Sunshine, and Number of passengers.

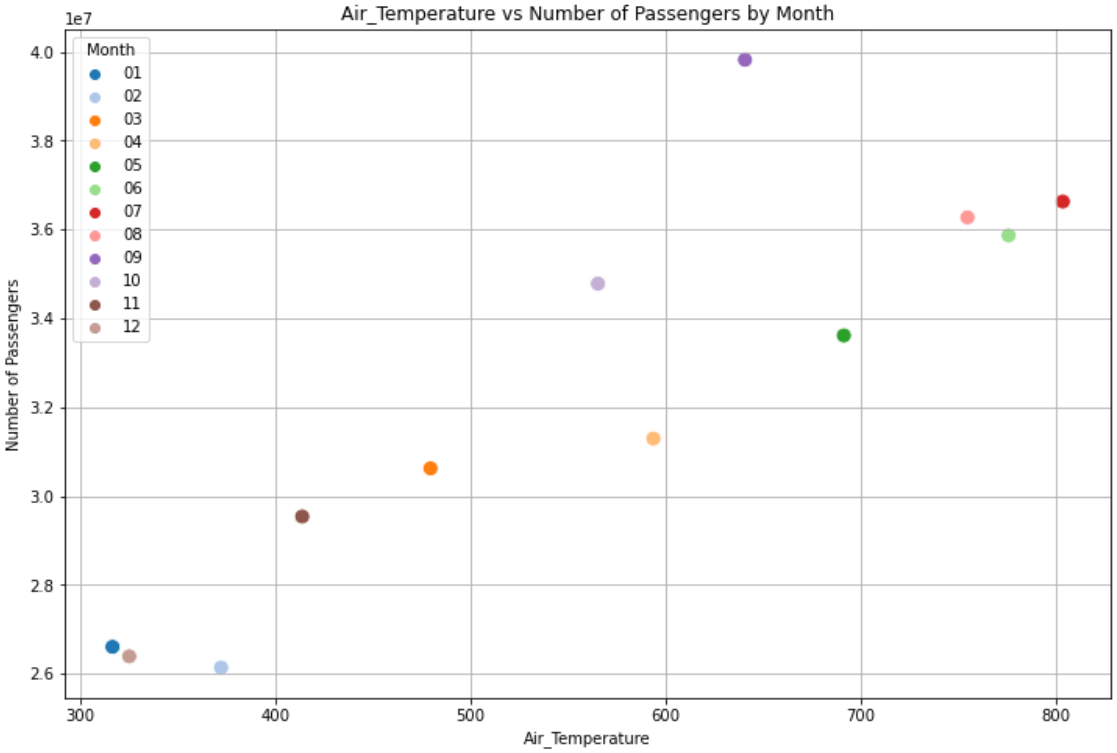
Creation of data pipeline and tests for the project

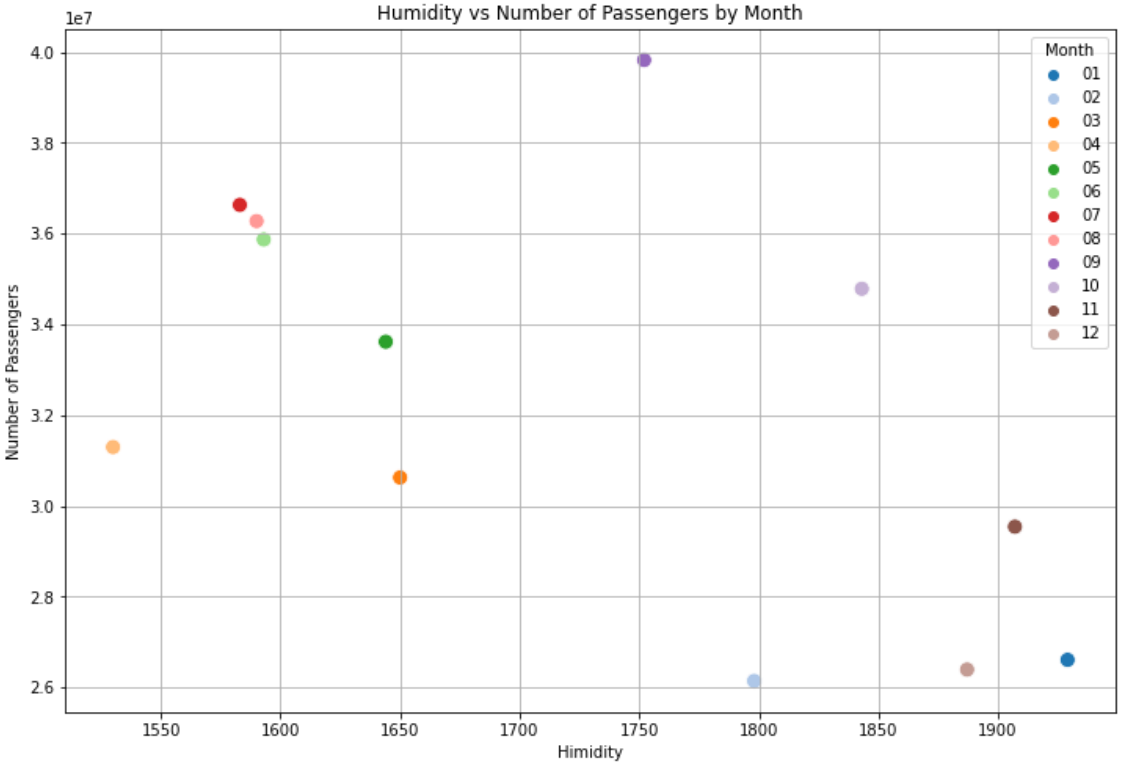
I set up an automated data pipeline for my project during this phase of the project by effectively following the project plan. I developed a Python script that effectively retrieves the chosen datasets from the internet, performs the required transformations, and fixes data issues. Following processing, the data is saved as SQLite databases in the /data directory. The pipeline.sh file, which is the pipeline's entry point, was placed in the same directory as the script, which was housed in the /project directory. When this bash script is run, the Python script is launched using the command python3 /project/pipeline.py, which starts the data pipeline.

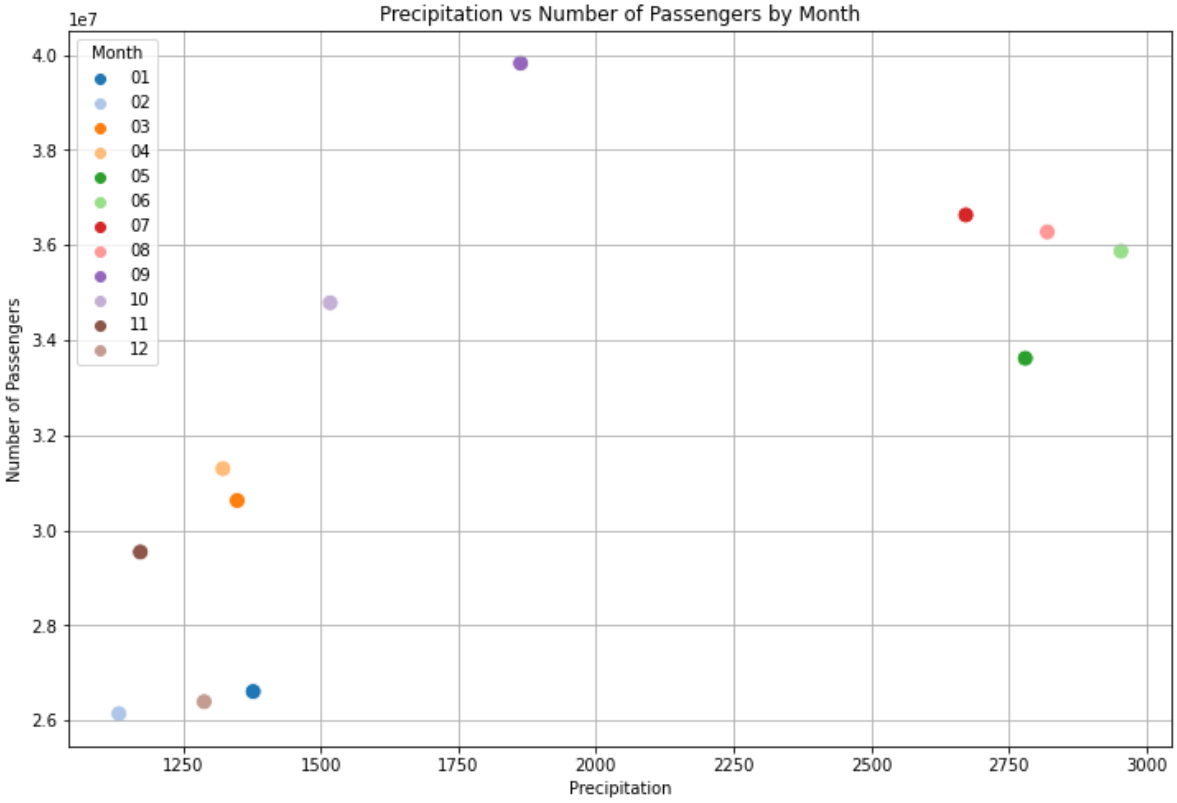
The script should result in datasets being present in the /data directory. It's crucial to remember that, following recommended practices, the real data sets were purposefully kept out of version control using.gitignore. I included automated tests in the form of a tests.sh file in the /project directory to ensure the data pipeline is reliable. A minimum of one system-test level scenario that verifies the data pipeline's execution and guarantees the production of the output file or files in the /data directory is one of the test cases that the tests.sh file is intended to run. It's noteworthy that version control does not verify these output files.

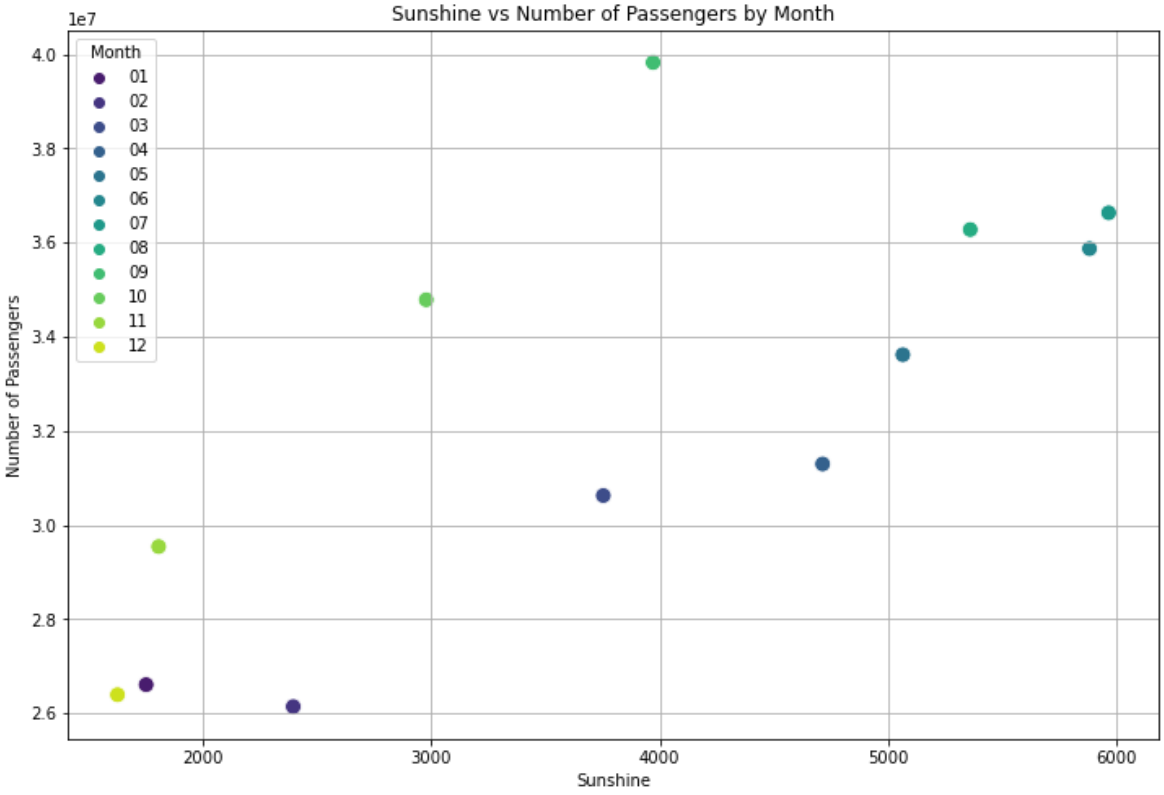
Results

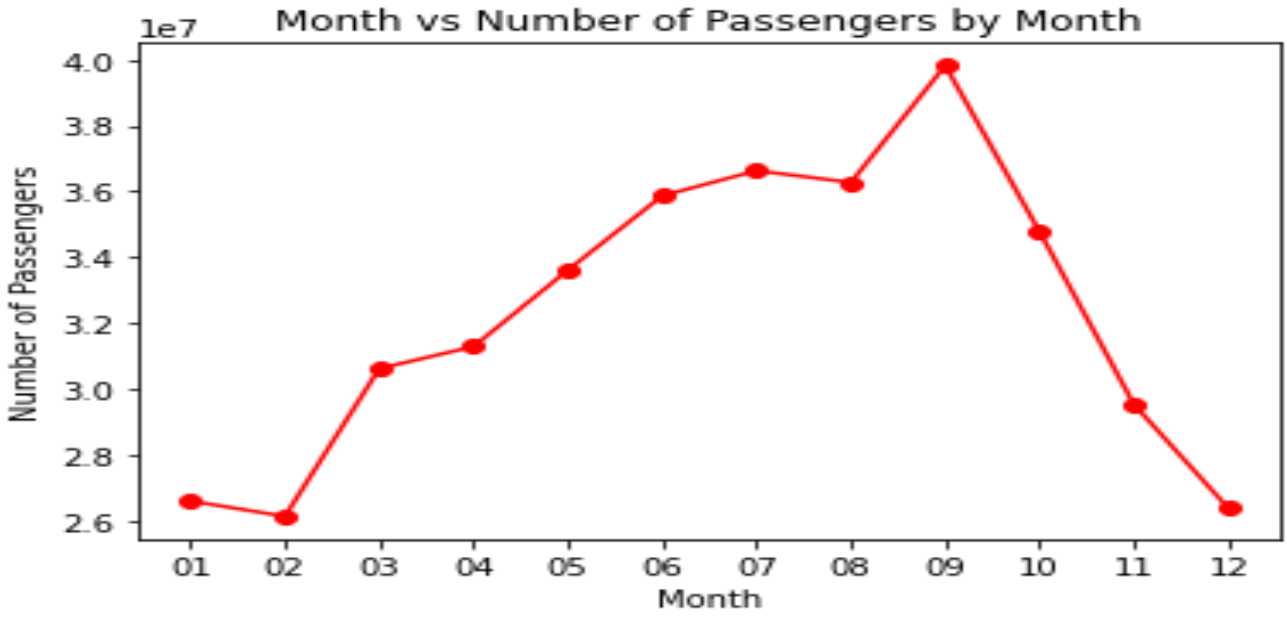


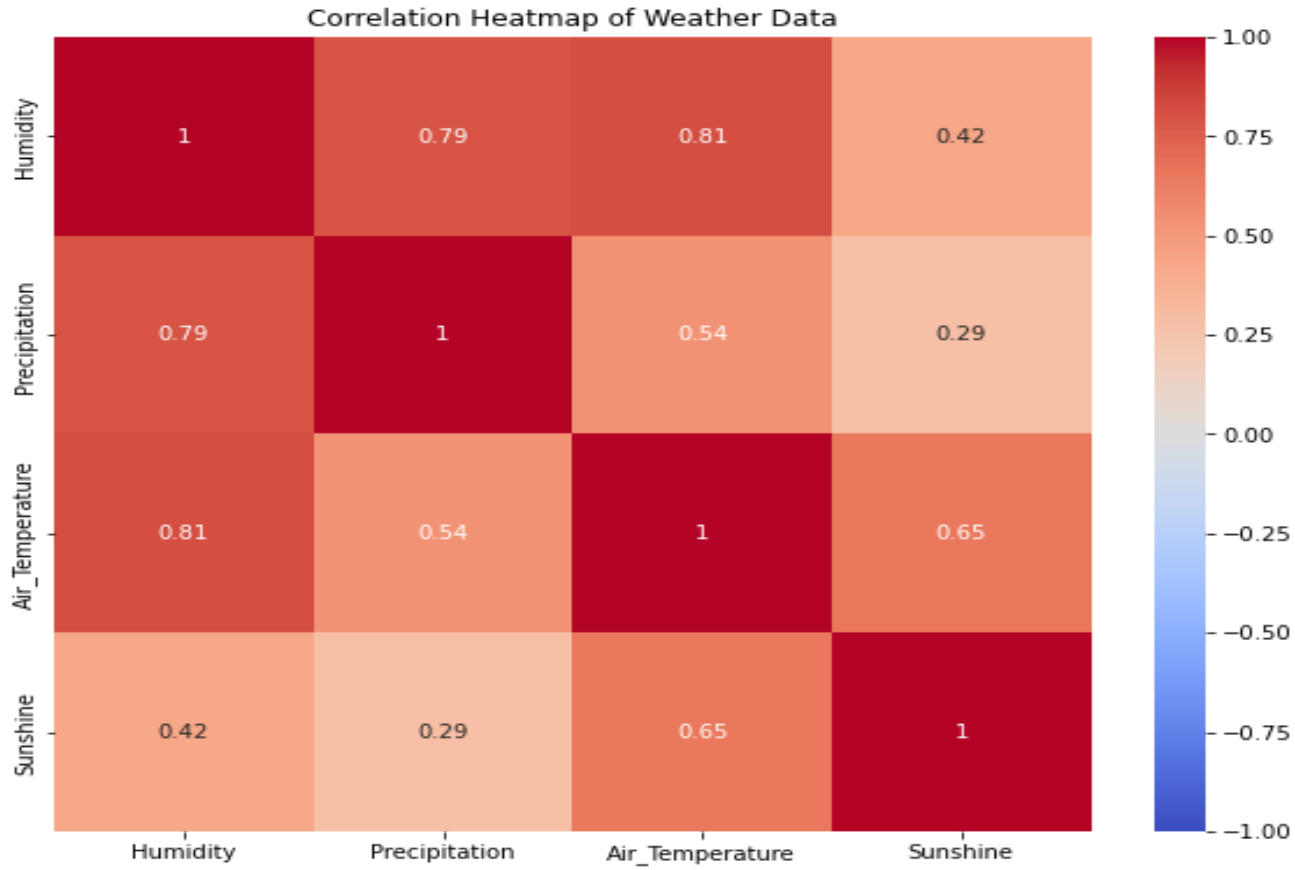

Air_Temperature vs Number of Passengers by Month
1e7
Number of Passengers
Air_Temperature
Month
01
02
03
04
05
06
07
08
09
10
11
12
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
300
400
500
600
700
800

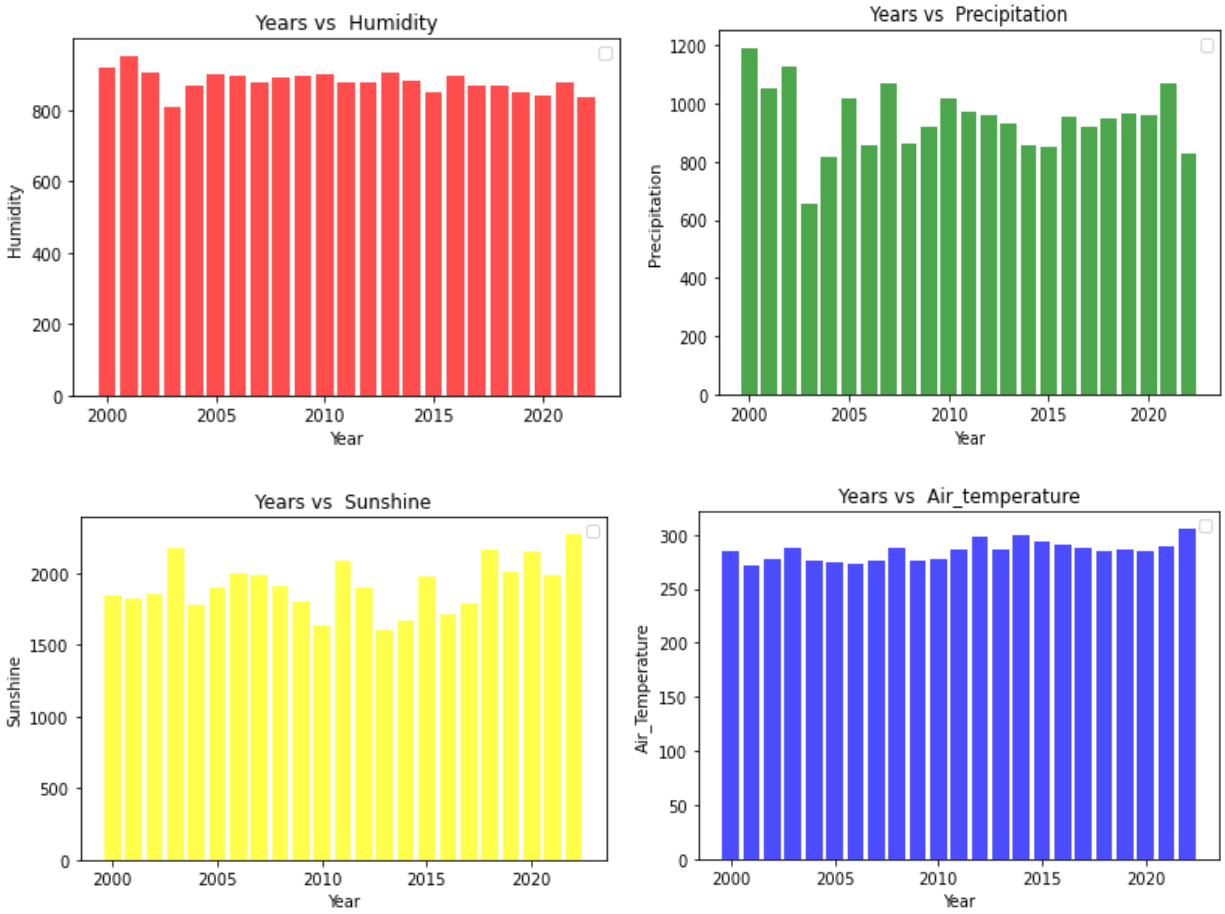

Humidity vs Number of Passengers by Month
1e7
Number of Passengers
Himidity
Month
01
02
03
04
05
06
07
08
09
10
11
12
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
1550
1600
1650
1700
1750
1800
1850
1900


Precipitation vs Number of Passengers by Month
1e7
Month
01
02
03
04
05
06
07
08
09
10
11
12
Number of Passengers
Precipitation
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
1250
1500
1750
2000
2250
2500
2750
3000


Sunshine vs Number of Passengers by Month
1e7
Month
01
02
03
04
05
06
07
08
09
10
11
12
Number of Passengers
Sunshine
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2000
3000
4000
5000
6000


Month vs Number of Passengers by Month
1e7
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
Number of Passengers
01
02
03
04
05
06
07
08
09
10
11
12
Month


Correlation Heatmap of Weather Data
Humidity
Precipitation
Air_Temperature
Sunshine
1
0.79
0.81
0.42
0.79
1
0.54
0.29
0.81
0.54
1
0.65
0.42
0.29
0.65
1
Humidity
Precipitation
Air_Temperature
Sunshine
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00



Analysis

1: Analysis of Number of Passengers Over the Years

The longitudinal examination of the number of passengers traveling reveals a consistent upward trend over the years, underscoring a positive trajectory in passenger volume. However, the unforeseen impact of the COVID-19 pandemic disrupted this growth pattern, resulting in a significant decline in travel numbers in 2020. The subsequent years, 2021, 2022, and 2023, depict a nuanced recovery scenario. While 2021 witnessed a marginal increase, 2022 and 2023 exhibited more substantial growth, signaling a resilient recovery after the pandemic-induced setback.

2: Analysis of Number of Passengers Across Different Months

Analyzing the monthly variations in passenger numbers indicates distinct travel patterns throughout the year. Notably, the month of September emerges as a peak travel period, marked by a notable surge in passenger volume. The observed trend illustrates a consistent rise from January to September, implying heightened travel activity during the earlier months. Conversely, a gradual decline is observed from September to December, indicative of a seasonal decrease in travel demand towards the end of the year. Understanding these seasonal variations is crucial for adapting operational strategies and resource allocation in response to shifting travel demands across different months.

3: correlation between different weather measures

We discovered that the temperature and humidity (0.79), temperature and sunshine (0.81), and humidity and sunshine (0.54) had the best positive connections. This indicates that there is a tendency for these variables to fluctuate in the same way. For instance, sunlight and humidity both tend to rise with temperature. Sunshine and precipitation have the most negative association (-0.75). This indicates that these factors often alter in opposition to one another. For instance, precipitation usually decreases with increased sunlight. Temperature and precipitation have relatively weak correlations (0.42) while humidity and precipitation have pretty poor correlations (0.29). This indicates that these factors have a somewhat positive association with one another. For instance, there is a little increase in precipitation along with higher temperatures. Overall, this heatmap demonstrates that temperature, humidity, and sunshine have high positive associations, whereas precipitation and sunshine have large negative relationships. Understanding how weather factors interact with one another and forecasting future weather conditions can both benefit from these relationships.

4: Analysis of how the weather is affecting the number of passengers traveling

They demonstrate that the quantity of people arriving at the airport is positively correlated with sunlight, humidity, and air temperature. This implies that an increase in sunlight, humidity, and air temperature corresponds to an increase in the number of people arriving at the airport. There is a negative relationship between precipitation and the volume of travelers arriving at the airport. This implies that fewer people usually fly into the airport during periods of increased precipitation.

Conclusion

The Analysis shows a strong relationship between the amount of passengers traveling and the weather. The epidemic caused passenger numbers to decline in 2020, but the following years—especially 2022 and 2023—showed a strong rebound. The monthly fluctuations revealed a significant increase in travel in September, with a steady increase preceding that month. Analyzing meteorological data, positive correlations between temperature, humidity, and sunlight highlight correlated patterns, whereas negative correlations between sunshine and precipitation highlight countercorrelated variations. These weather dynamics have a major role in passenger travel; greater air temperature, humidity, and sunshine all favorably correlate with passenger arrivals, whereas higher precipitation levels are negatively correlated with arrivals. This complex interaction highlights the necessity for flexible transportation plans that take climatic conditions into account and use them to their advantage to maximize operational effectiveness and forecast passenger demand