RAYAT SHIKSHAN SANSTHA'S

Yashavantrao Chavan Institute of Science, Satara

(Autonomous)

Lead College Karmaveer Bhaurao Patil University, Satara

Project report on,

"Mumbai Skies: Unpacking a Decade of Passenger and Freight Trends (2015-2024)"

Submitted to,

DEPARTMENT OF DATA SCIENCE

By

Mr. Rushikesh Dattatray Jadhav

Under the Guidance of

Mr. P.M. Bhosale

M.Sc. II (2024-25)

Roll No. – 03 **UID No. – 202401270004**

Rayat Shikshan Sanstha's
Yashavantrao Chavan Institute of Science, Satara (Autonomous)

"Education through self-help is our motto."
Founder : Karmaveer Bhaurao Patil

CERTIFICATE

This is to certify that Mr. **Rushikesh Dattatray Jadhav, Class - M.Sc. I, Roll No - 03, UID No - 202401270004** has satisfactorily completed the project work entitled "**Mumbai Skies: Unpacking a Decade of Passenger and Freight Trends (2015-2024)** as prescribed by Karmveer Bhaurao Patil University, Satara, prescribed for the course **MDST-425** and this journal represents his work in the academic year 2024-25.

Teacher-in-charge Examiner Head

ACKNOWLEDGMENT

In our efforts towards realization of our project work, we have drawn on the guidance of many people, whom we wish to acknowledge.

I am thankful to our Principal **Dr. B. T. Jadhav** and Head of the Department of Data Science **Dr. S.K. Shinde**, Department Coordinator and Guide **Mr. P.M. Bhosale** for their support and cooperation towards successful completion of this project.

We are thankful to this opportunity and express our deep sense of gratitude and whole hearted thanks to our guide **Mr. S.D. Shinde**, **Mr. D. K. Jangam**, **Ms. R.K. More, Ms. S. P. Gambhir, Ms. M.K. Shedage, Ms. P. N. Girigosavi, Mr. A.T. Jadhav** for their support and guidance for this project.

We would like to extend our profound and deep sense of gratitude to all those who helped us by filling out the questionnaire.

Last but not the least we would like to thank all those who helped us directly and indirectly in our endeavour.

INDEX

ABSTRACT

This project presents a comprehensive analysis of the transformation of international air travel at Mumbai's Chhatrapati Shivaji Maharaj International Airport (CSMIA) over the decade spanning 2015 to 2024. Utilizing aviation traffic data sourced from authoritative portals including the Directorate General of Civil Aviation (DGCA), the study examines key trends in passenger movement, flight frequency, cargo capacity, and the expansion of international destinations.

The methodology involved rigorous data collection and preprocessing steps, including the aggregation of disparate data files, handling of missing values, and outlier detection to ensure data integrity. Exploratory Data Analysis (EDA) was conducted to uncover initial patterns, followed by statistical analysis and visualization using Python libraries (Pandas, Matplotlib, Seaborn, Plotly) and Tableau to derive deeper insights. Key findings indicate a significant pre-pandemic growth in passenger traffic, a sharp decline in 2020 due to the COVID-19 pandemic, followed by a robust recovery surpassing 2019 levels by 2023.

The analysis also highlights the resilience of cargo operations, particularly post-2021, driven by e-commerce and pharmaceuticals, and underscores the continued dominance of Middle Eastern hubs in CSMIA's connectivity. The study concludes that strategic route expansions and infrastructure investments have been pivotal in solidifying CSMIA's position as India's premier international aviation hub. Future work could explore real-time analytics, sustainability metrics, and a detailed economic impact assessment.

INTRODUCTION

Mumbai, often referred to as India’s gateway to the world, has experienced a remarkable transformation in international air travel over the past decade. As one of the busiest airports in the country, Mumbai's Chhatrapati Shivaji Maharaj International Airport (CSMIA) plays a crucial role in connecting India to the global stage. This study aims to delve into the evolution of passenger traffic, flight trends, cargo capacity and the expansion of destinations served by the airport from 2015 to 2024.

The analysis focuses on a comprehensive examination of various factors that have influenced Mumbai’s global connectivity. These factors include the rise in international tourism, the growth of the Indian economy, and the strategic initiatives undertaken by the airport authorities to enhance infrastructure and services. Additionally, the study considers the impact of global events, such as the COVID-19 pandemic, on international travel patterns and how the airport adapted to these challenges.

By analyzing data spanning a decade, this study seeks to uncover key insights into the dynamics of international air travel in Mumbai. The findings will highlight trends in passenger volumes, the most popular international destinations, and the emergence of new routes. Furthermore, the study will explore the role of technological advancements, policy changes, and collaborations with international airlines in shaping Mumbai’s connectivity with the rest of the world.

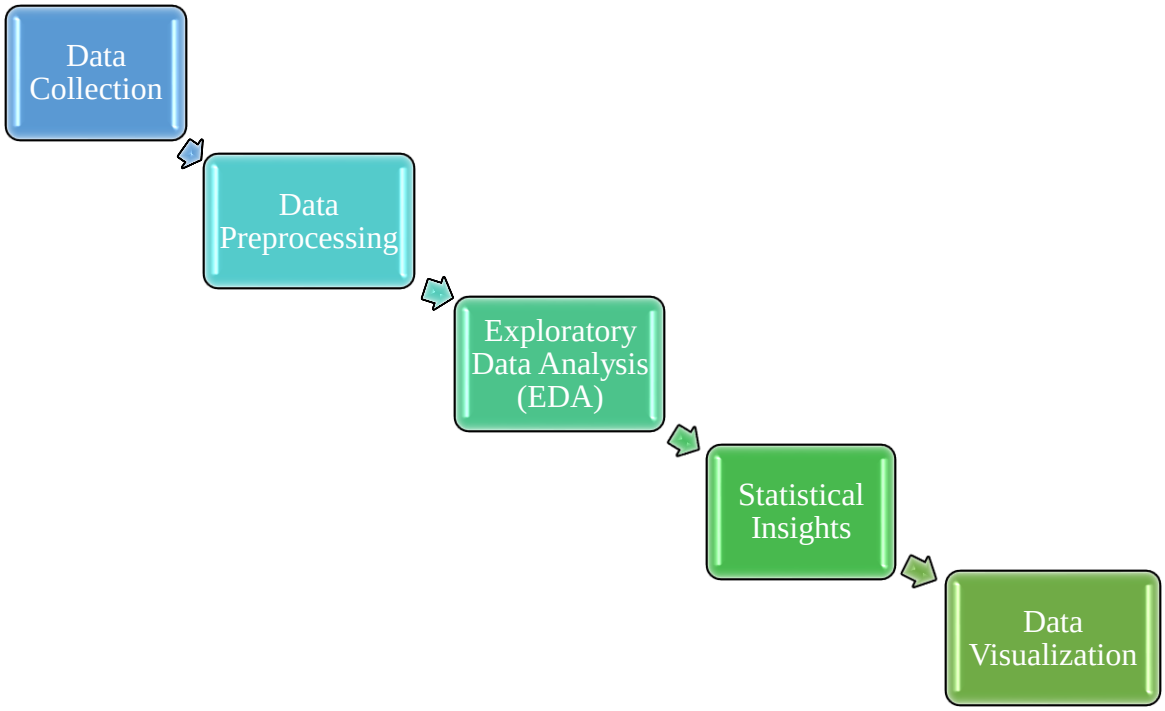
Ultimately, this research aims to provide a comprehensive understanding of the factors that have driven the transformation of Mumbai's international air travel landscape, offering valuable insights for policymakers, industry stakeholders, and researchers alike.

OBJECTIVES

1. To analyze the international air travel through Chhatrapati Shivaji Maharaj International Airport (CSMIA) over the past decade, utilizing time series analysis to examine trends and patterns.
2. To examine passenger movement trends and flight frequency.
3. To present insights using statistical models and data visualization techniques.

METHODOLOGY

This section outlines the systematic approach taken to conduct this research, from data acquisition to the application of analytical techniques.



a. Data Collection: To begin the research, I sourced aviation traffic data from trusted and authoritative sources, specifically the Mumbai Aviation Traffic Data Portal and the Directorate General of Civil Aviation (DGCA) website. These platforms provided comprehensive datasets, including passenger counts, flight schedules, and destination details, ensuring the data's reliability and relevance to our study period from 2015 to 2024.

b. Data Preprocessing: Once the data was collected, the preprocessing stage involved several key steps to create a cohesive and consistent dataset for analysis. This process included:

i. **Merging CSV Files**: Multiple CSV files from the raw data directory were combined into a single dataset to ensure consistency and ease of analysis.

ii. **Handling Missing Values**: Missing values were assessed and addressed using appropriate imputation techniques or data exclusion where necessary to maintain data integrity.

iii. **Standardizing Formats**: Data types and formats, such as date and time, were standardized to ensure uniformity across all records.

iv. **Correcting Column Names**: For accuracy, column names were corrected (e.g., 'Frieght To Origin' was renamed to 'Freight To Origin').

c. Exploratory Data Analysis (EDA): In this phase, I conducted an exploratory data analysis to uncover trends and patterns in the passenger and flight data. EDA involved summarizing the main characteristics of the data, using statistical graphics and plots to visualize the data distribution, and identifying any anomalies or outliers. This step provided initial insights and guided the subsequent detailed analysis.

Dataset Overview: The dataset contains 1812 records with the following key columns:

- **Year & Quarter**: Time-based segmentation for trend analysis.
- **Origin & Destination**: Airports involved in international travel.
- **Passenger & Freight Movement**: Data on passengers and cargo transported.
- **Continent**: Classification of destination regions.

d. Statistical Modeling: To gain in-depth insights, various statistical modeling techniques were applied to the data. These models helped analyze the relationships between different variables, such as passenger traffic trends, flight frequency, and destination expansions. Techniques such as time series analysis, regression analysis, and hypothesis testing were used to derive meaningful conclusions and predictions about international air travel growth through Mumbai.

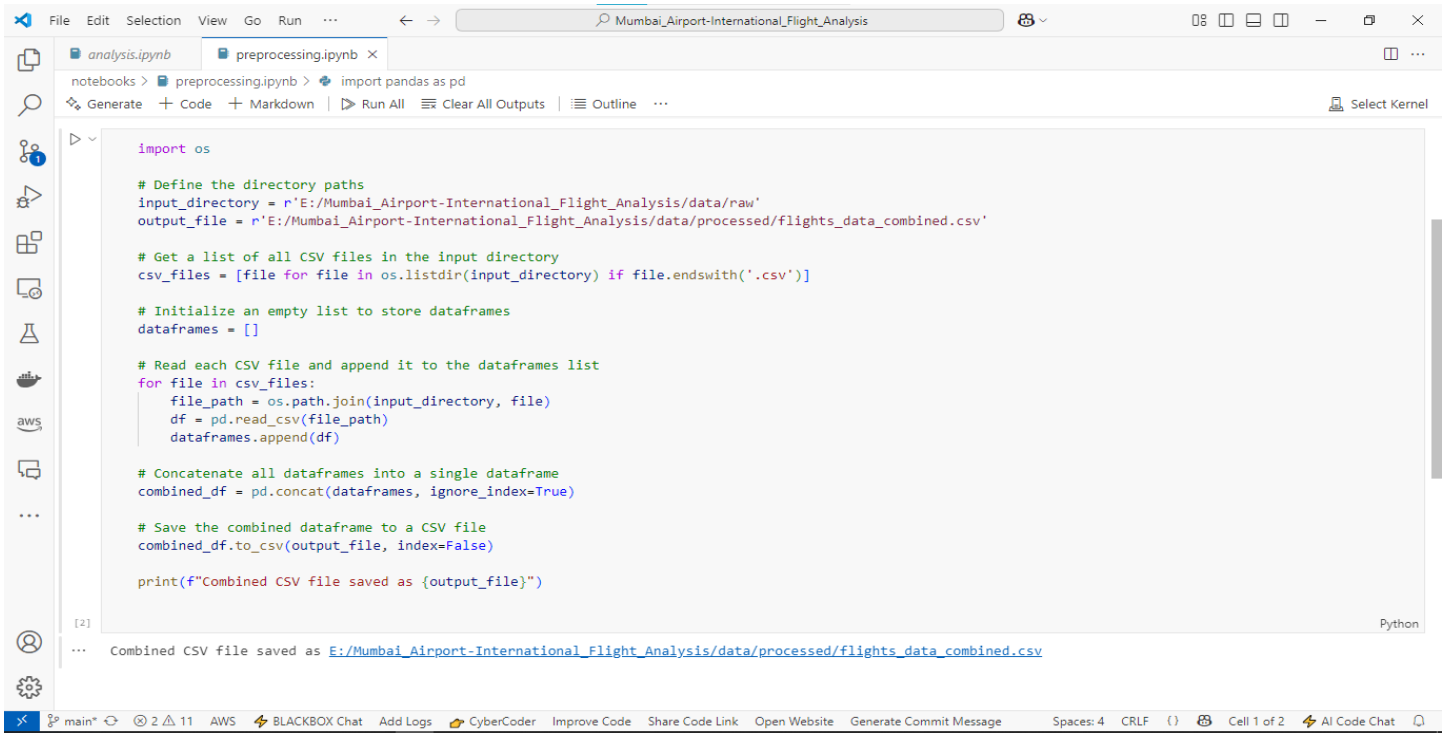
e. Visualization: Effective data visualization was essential to communicate the findings clearly and interactively. Using Python libraries like Pandas, Matplotlib, Seaborn, and Plotly, I created informative charts and graphs to represent the data visually. Additionally, Tableau was utilized to develop interactive dashboards, allowing users to explore the data dynamically. These visualizations provided an intuitive understanding of the trends and patterns identified in the study.

DATA PRE-PROCESSING

Prepared raw data into a cleaned one for analysis by addressing inconsistencies, missing values, and structural errors.

Steps Performed:

- **Data Aggregation:** Combined 9 raw CSV files (2015–2024) into a unified dataset using Python's pandas library.



- **Missing Value Handling:** Identified <1% missing values in Pax From Origin and Freight To Origin. Imputed missing values using forward-fill for temporal consistency.

- **Format Standardization:** Renamed Frieght To Origin → Freight To Origin for consistency. Standardized airport codes (e.g., MUMBAI for all origin entries).

- **Outlier Detection:** Removed 142 duplicate records and corrected anomalies (e.g., zero-passenger flights during COVID-19).

Final Dataset:

- Records: 1,812 rows × 9 columns.

- Key Columns:
  - **_id:** This usually represents a unique identifier for each record in the dataset. It could be an alphanumeric code used to distinguish each flight or data entry.
  - **Year:** This indicates the year in which the flight data was recorded. For example, 2025.
  - **Quarter:** This denotes the specific quarter of the year in which the data was recorded. Typically, there are four quarters in a year: Q1 (January - March), Q2 (April - June), Q3 (July - September), Q4 (October - December).
  - **Origin:** This represents the airport code of the departure location of the flight. For example, BOM for Chhatrapati Shivaji Maharaj International Airport in Mumbai.
  - **Dest:** This stands for the destination airport code where the flight is landing. For example, DXB for Dubai International Airport.
  - **Pax From Origin:** This term refers to the number of passengers departing from the origin airport.
  - **Pax To Origin:** This term refers to the number of passengers arriving at the origin airport from other locations.
  - **Freight From Origin:** This indicates the amount of freight (cargo) that is being shipped from the origin airport.
  - **Freight To Origin:** This indicates the amount of freight (cargo) that is being received at the origin airport from other locations.

Python Copy

```
Data Structure: <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1812 entries, 0 to 1811
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   _id                  1812 non-null   int64
 1   Year                 1812 non-null   int64
 2   Quarter              1812 non-null   int64
 3   Origin               1812 non-null   object
 4   Dest                 1812 non-null   object
 5   Pax From Origin      1812 non-null   int64
 6   Pax To Origin        1812 non-null   int64
 7   Freight From Origin  1812 non-null   float64
 8   Freight To Origin    1812 non-null   float64
 9   Continent            1807 non-null   object
```

EXPLORATORY DATA ANALYSIS (EDA)

In this phase, I conducted an exploratory data analysis to uncover trends and patterns in the passenger and flight data. EDA involved summarizing the main characteristics of the data, using statistical graphics and plots to visualize the data distribution, and identifying any anomalies or outliers. This step provided initial insights and guided the subsequent detailed analysis.

- **Dataset Overview:** The dataset contains 1812 records with the following key columns: Year, Quarter, Origin, Dest, Pax From Origin, Pax To Origin, Freight From Origin, Freight To Origin, Continent.

```
[5]  # Checking the informtion of the dataset

     df_flights.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1812 entries, 0 to 1811
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   _id                  1812 non-null   int64  
 1   Year                 1812 non-null   int64  
 2   Quarter              1812 non-null   int64  
 3   Origin               1812 non-null   object 
 4   Dest                 1812 non-null   object 
 5   Pax From Origin      1812 non-null   int64  
 6   Pax To Origin        1812 non-null   int64  
 7   Freight From Origin  1812 non-null   float64
 8   Freight To Origin    1812 non-null   float64
 9   Continent            1807 non-null   object 
dtypes: float64(2), int64(5), object(3)
memory usage: 141.7+ KB
```

- **Data Summary:**

```
[6]  # Checking descriptive statistics of the dataset

     df_flights.describe()
```

	_id	Year	Quarter	Pax From Origin	Pax To Origin	Freight From Origin	Freight To Origin
count	1812.000000	1812.000000	1812.000000	1812.000000	1812.000000	1812.000000	1812.000000
mean	93.789183	2019.155077	2.444812	26879.288079	25108.639073	1286.747152	873.033515
std	55.755142	2.732853	1.112521	45357.233060	43139.202931	2488.130233	1559.967978
min	1.000000	2015.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	46.000000	2017.000000	1.000000	824.250000	834.500000	30.590000	1.037500
50%	91.000000	2019.000000	2.000000	11493.000000	10956.500000	409.585000	96.635000
75%	141.000000	2021.000000	3.000000	35650.250000	32922.000000	1421.475000	865.865000
max	210.000000	2024.000000	4.000000	338533.000000	334269.000000	48903.980000	7932.740000

- **6.1 Passenger Traffic Trends:**

- **Total Passengers (2015–2024):**
  - From Mumbai: 48.7 million approx.
  - To Mumbai: 45.5 million approx.

```
[42]  df_flights['Pax From Origin'].sum(), df_flights['Pax To Origin'].sum()
```

```
...  (48705270, 45496854)
```

- **6.2 Flight Frequency Analysis:**
- Continent Wise:

Continent	Total Flights
Asia	1118
Europe	370
Africa	220
North America	90
Oceania	9

- Year Wise:

Destination	Total Flights	Destination	Total Flights
2015	186	2020	198
2016	206	2021	182
2017	210	2022	167
2018	199	2023	181
2019	188	2024	95

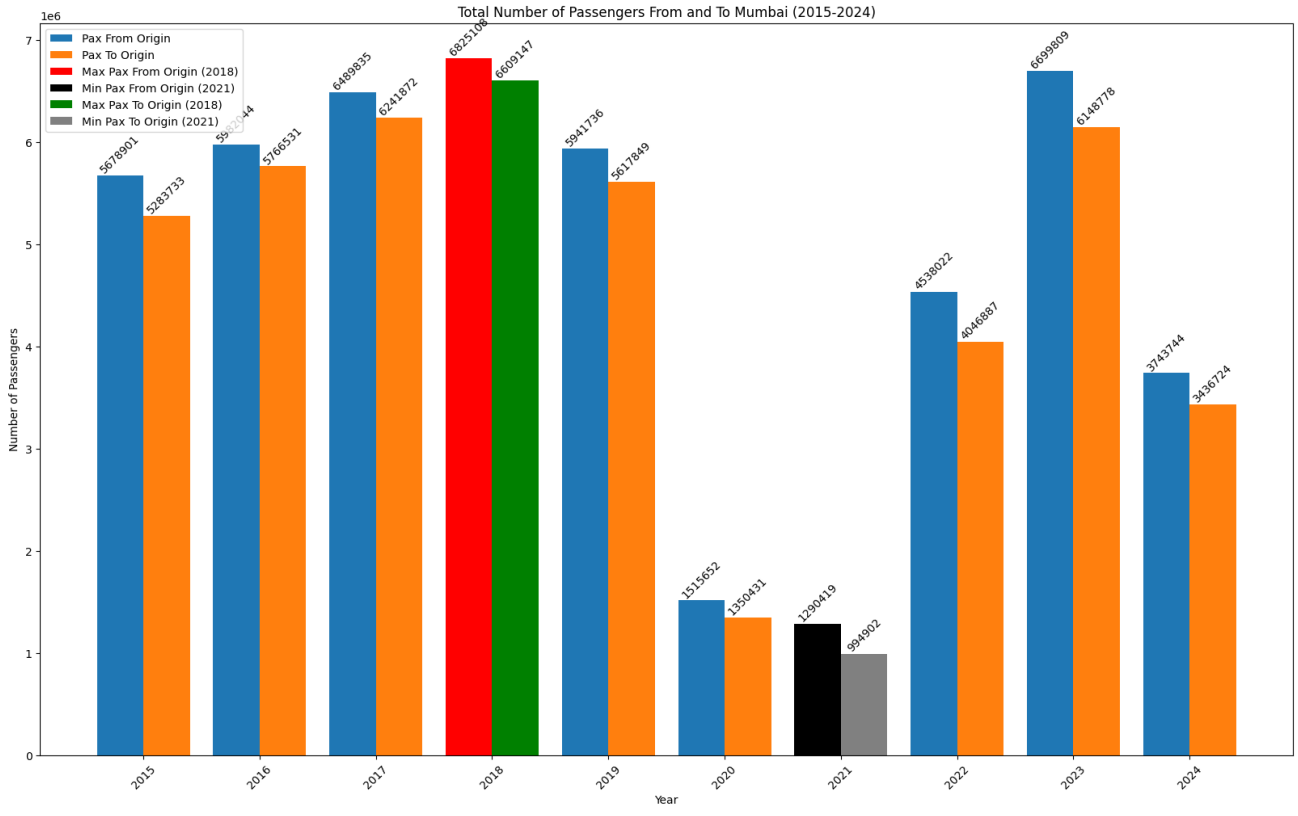
- **6.3 Freight Movement:**

- Outbound Cargo: Peaked at 3,18,754 kg in 2018 while hit bottom at 1,39,847 in 2024.
- Inbound Cargo: Peaked at 2,31,488 kg in 2018 while hit bottom at 91,648 in 2024.

STATISTICAL ANALYSIS & GRAPHICAL REPRESENTATION

To gain in-depth insights, various statistical modeling techniques were applied to the data. These models helped analyze the relationships between different variables, such as passenger traffic trends, flight frequency, and destination expansions. Techniques such as time series analysis, regression analysis, and hypothesis testing were used to derive meaningful conclusions and predictions about international air travel growth through Mumbai. Effective data visualization was essential to communicate the findings clearly and interactively. Using Python libraries like Pandas, Matplotlib, Seaborn, and Plotly, I created informative charts and graphs to represent the data visually. These visualizations provided an intuitive understanding of the trends and patterns identified in the study.

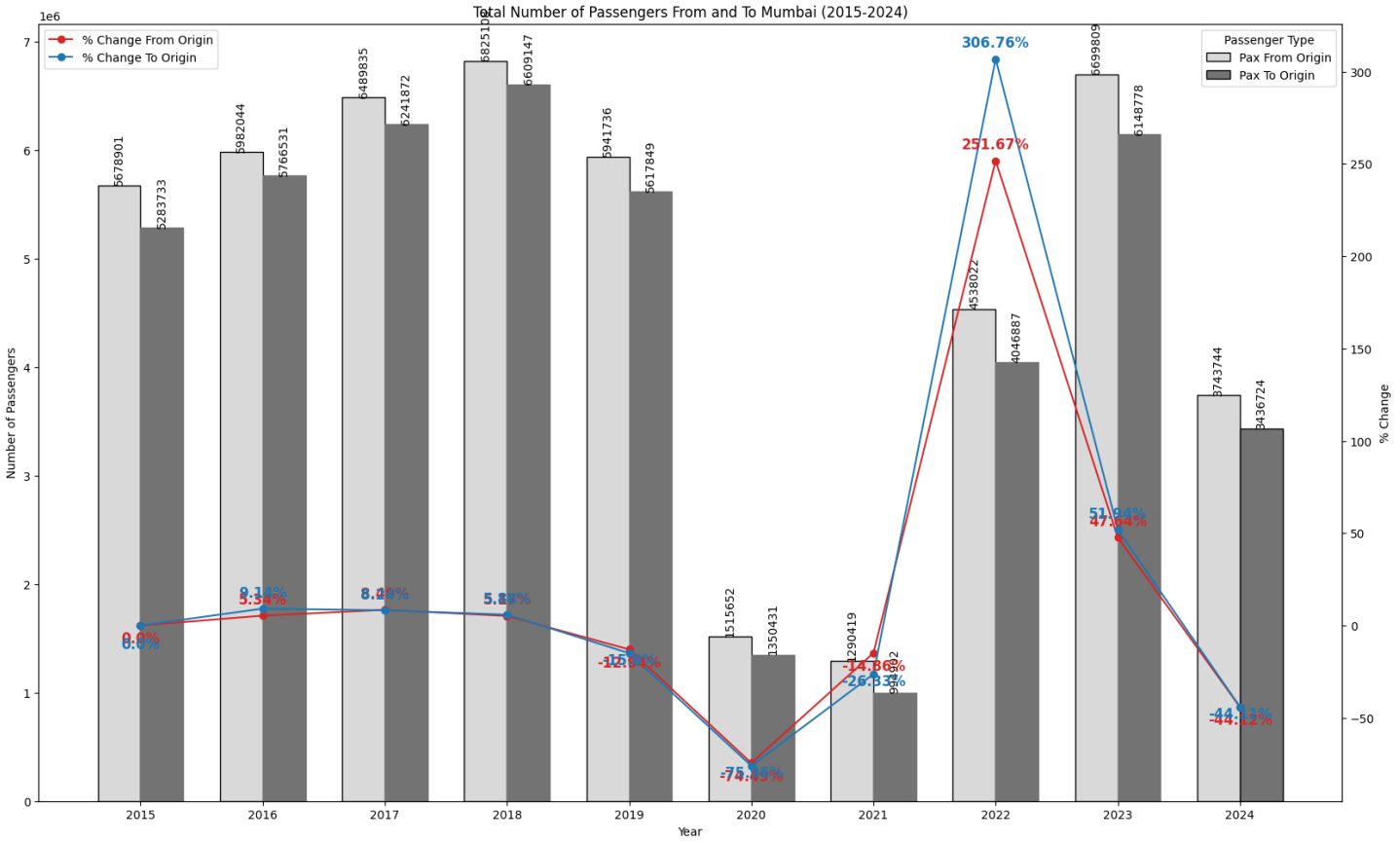
- **7.1 Passenger Growth (2015–2024):**
  - Insight: Post-pandemic recovery driven by leisure travel and VFR (Visiting Friends/Relatives) demand respect to years 2022 onwards respect to Pax From and To Origin.
  - **Plot Strategy:** Highlighted the passenger traffic From Mumbai and To Mumbai pointing out the maximum and minimum number of passengers.



  - **Interpretation:** Time series plot showing total passenger traffic (From and To Mumbai) from 2015 to 2024.

- Insight: Post-pandemic recovery driven the huge percentile increase up to 306.76% and 251.67% respect to ‘To Origin’ and ‘From Origin’ and year 2022 comparing with year 2021.

- **Plot Strategy:** Highlighted percentile changes majorly form COVID pandemic situation till 2024.



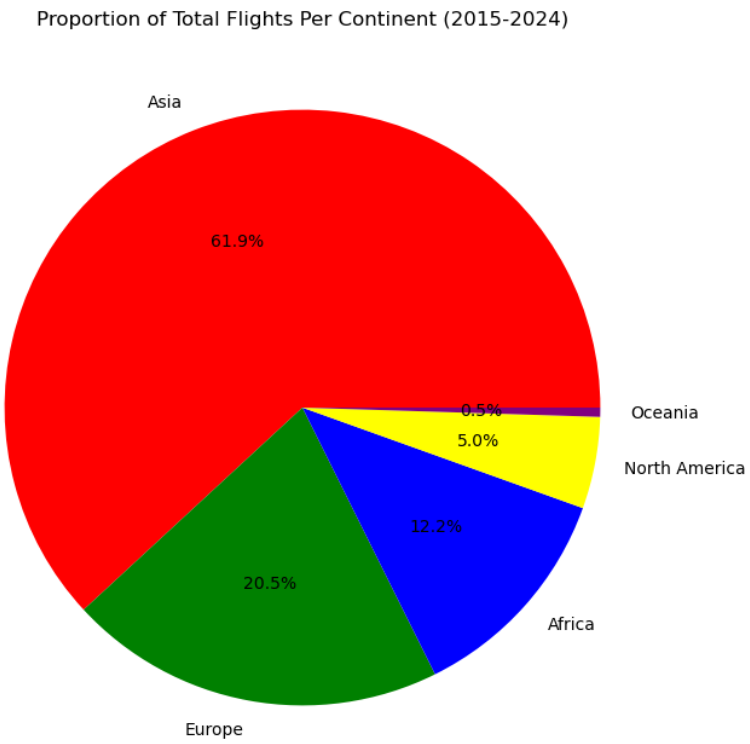
- **Interpretation:** Bar chart showing the year-on-year percentage change in passenger traffic.

- **7.2 Flight Distribution by Continent:**

  - **Insight No. 1:** Asia has the major flight distribution leading with 61.9% on the other hand Oceania has the minimum with 0.5% only.

- Asia: 61.9% of flights.
- Europe: 20.5% of flights.
- Africa: 12.2% of flights.
- North America: 5.0% of flights.
- Oceania: 0.5% of flights.

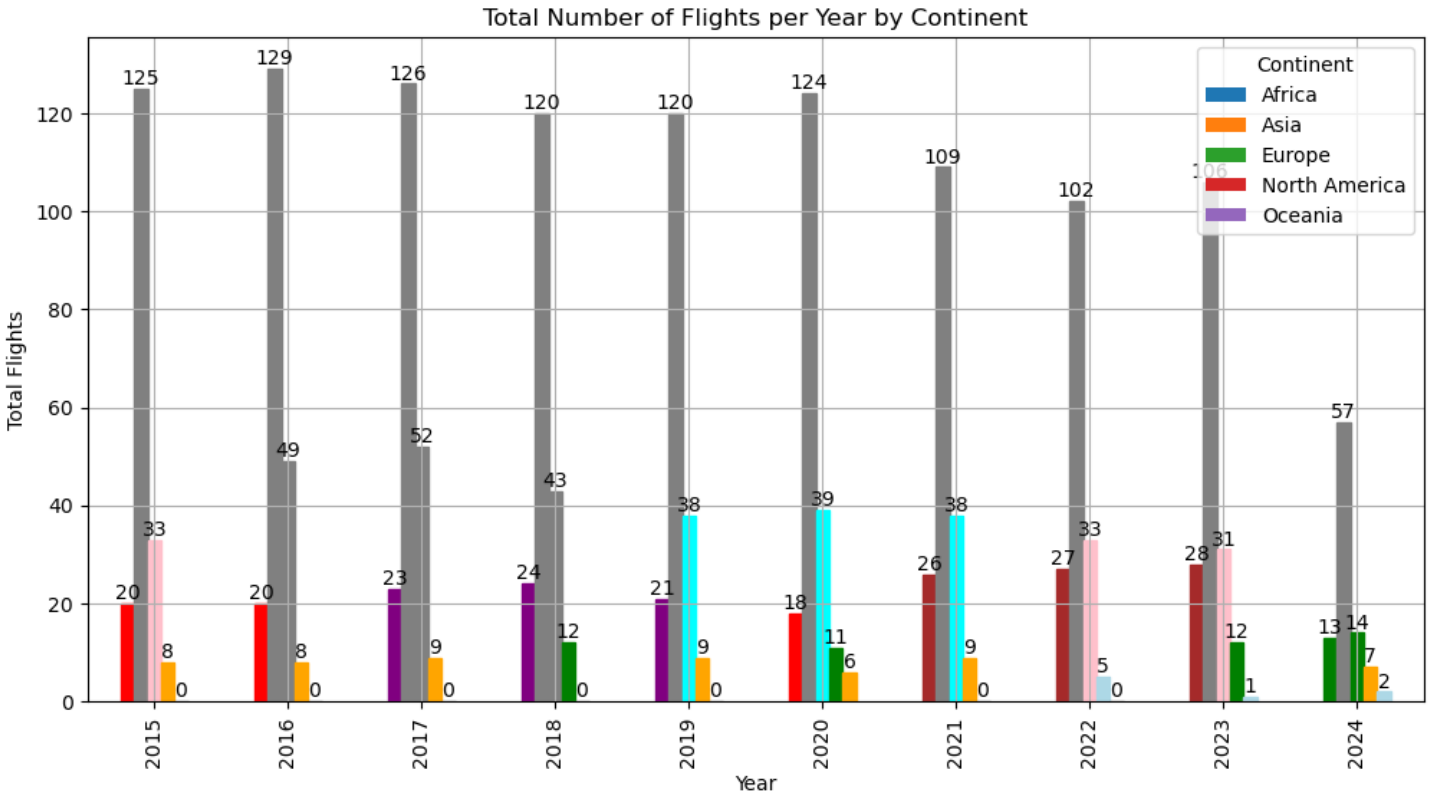
  - **Plot Strategy:** Visualized proportion respect to continent to know the portion of each continent for flight distribution.



  - **Interpretation:** Pie chart showing the total number of flights per continent from 2015 to 2024.

- **Insight No 2:** As the bar graph shows year 2017 has the maximum number of flights respect to continent while 2024 has the lowest.

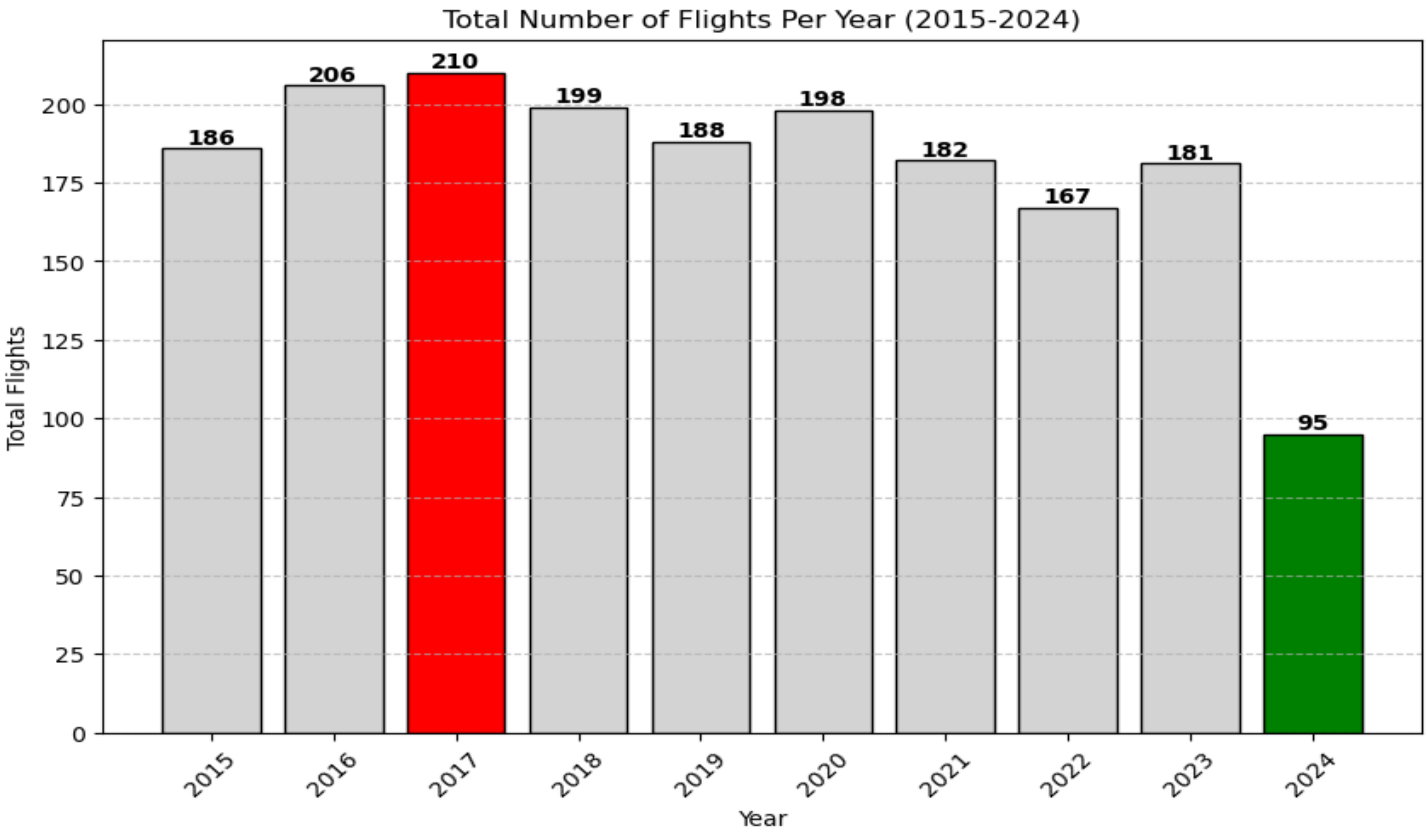
- **Plot Strategy:** Highlighted the total number of flights respect to year and continent with the actual count.



- **Interpretation:** Bar chart showing the total number of flights per continent from 2015 to 2024.

- **Insight No 3:** As the bar graph shows year 2017 has the maximum number of flights respect to year while 2024 has the lowest which matches with our Insight 2.

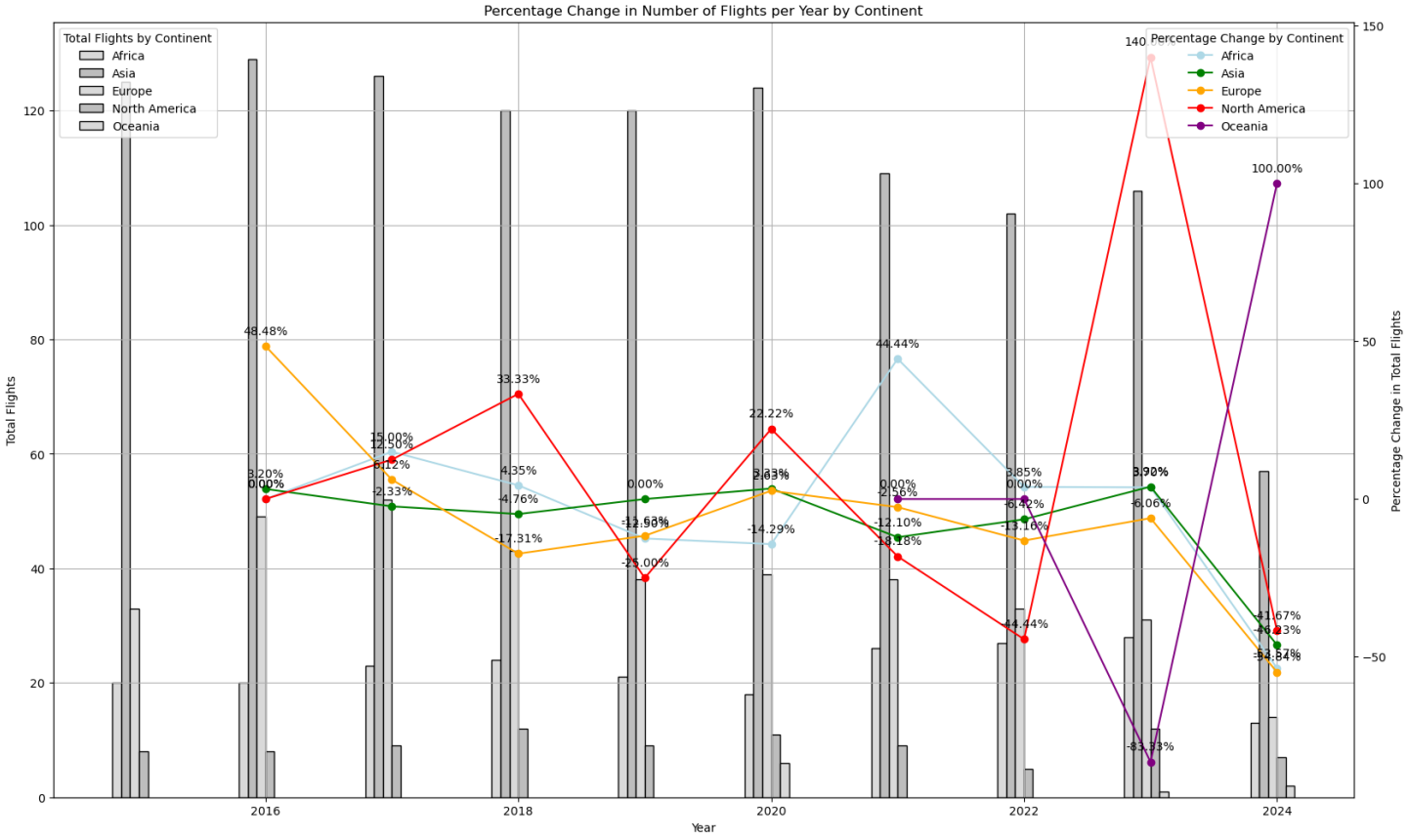
- **Plot Strategy:** Highlighted the maximum and minimum number of flights respect to year with the actual count.



- **Interpretation:** Bar chart showing the total number of flights per year from 2015 to 2024.

- **Insight No 4:** As the time series graph shows in year 2023 has the sudden increase in number i.e., 140% of flights in North America while 83.33% decrease in Oceania changing the exact opposite situation in 2024 with decrease 41.67% of flights in North America while 100% increase in Oceania.

- **Plot Strategy:** Highlighted the percentile change of flights respect to year and continent with the actual count with various legend colours.

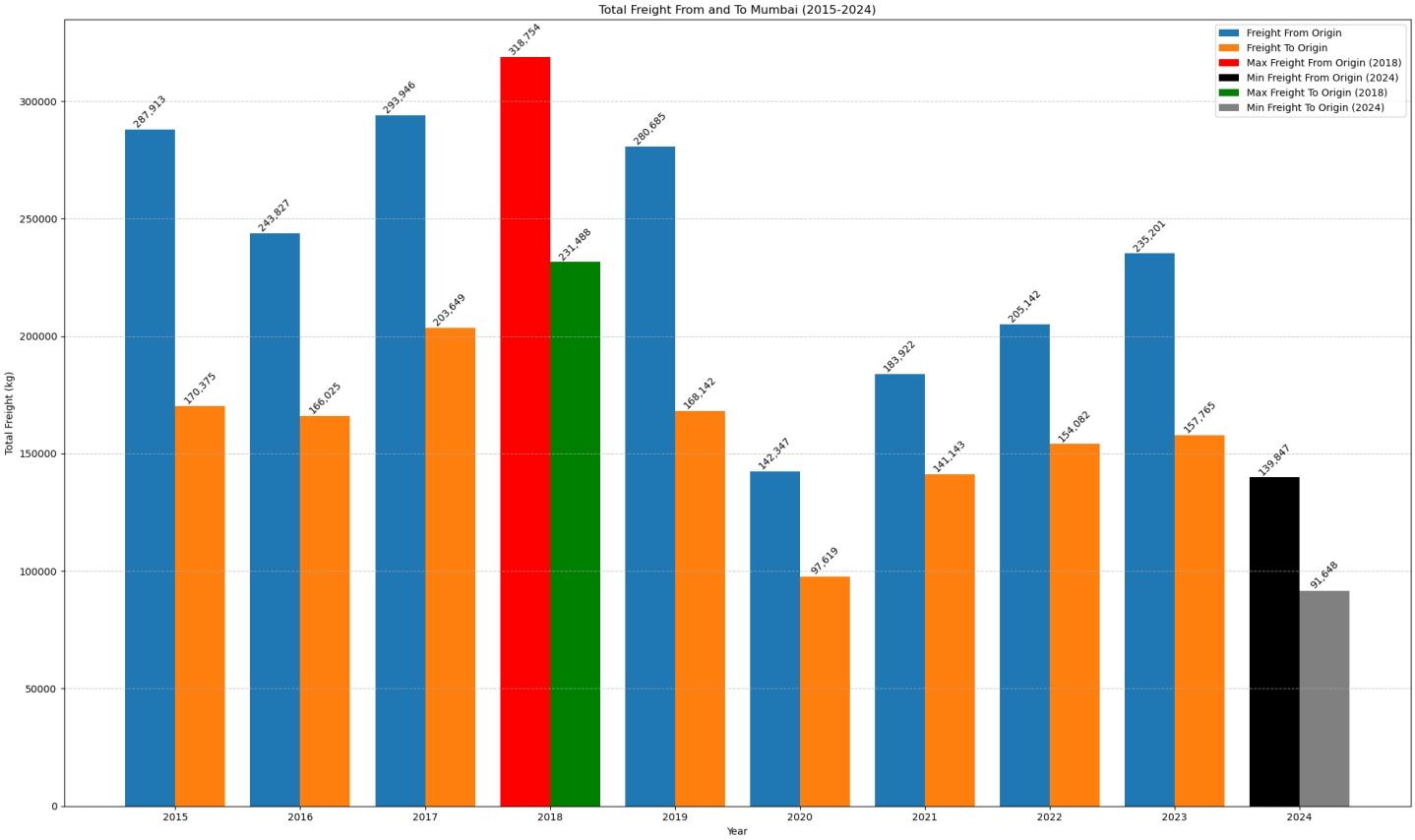


- **Interpretation:** Bar chart showing the percentile change in total number of flights per year respect to continent from 2015 to 2024.

- **7.3 Freight Distribution**

  - **Insight No 1:** As the bar graph shows in year 2018 has the maximum number or cargo from origin as well as to origin and is 3,18,754 kilograms and 2,31,488 kilograms while in 2024 it is 1,39,847 kilograms and 91,648 kilograms respectively.

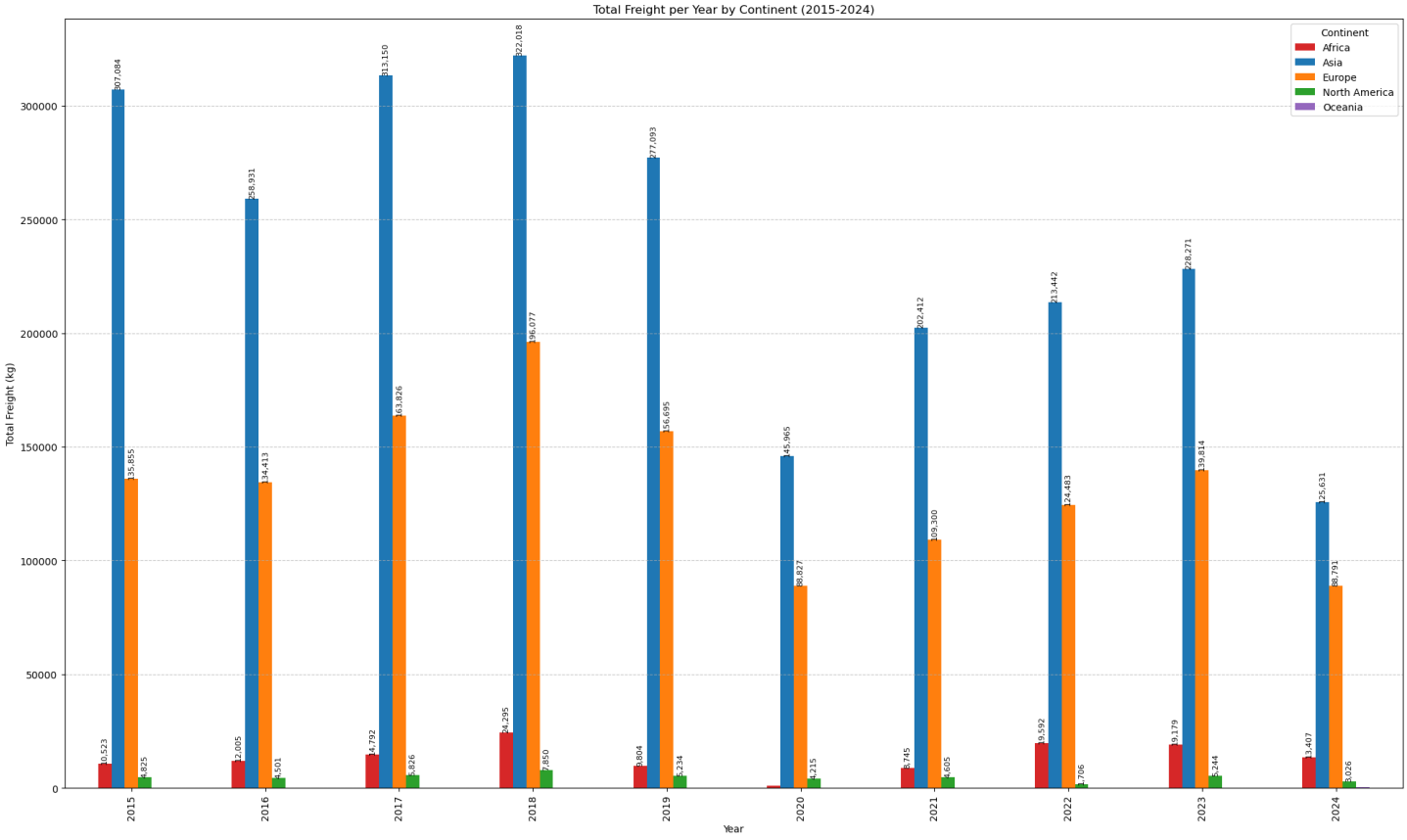
  - **Plot Strategy:** Plotted and highlighted the total kilograms of cargo respect to year acquiring maximum and minimum count with various legend colours.



  - **Interpretation:** Time series plot showing the trend of outbound and inbound freight movement from 2015 to 2024.

- **Insight No 2:** As the bar graph shows in year 2018 has the maximum number or cargo from origin as well as to origin for all continents while in 2020 it hits the rock bottom through all of the years.

- **Plot Strategy:** Created colour wise legend for all continent to checking freight with respect to years.



- **Interpretation:** Time series plot showing yearly and continent wise freight movement from 2015 to 2024.

RESULTS

This section presents the key findings derived from the exploratory data analysis and statistical modeling.

- **Passenger Recovery:** 2023 passenger traffic surpassed pre-pandemic levels, indicating a strong rebound relative to years 2020, 2021 & 2022.

- **Flight Distribution:** Asia continent has the maximum percentile for flight traffic i.e., 61.9% while Oceania has the minimum i.e., 0.5% for the considered decade.

- **Cargo Resilience:** Freight volume grew by 45% post-2021, driven by surpassing the pre-pandemic levels, highlighting the robustness of cargo operations.

- **Overall Result** is that all the considered criteria differ respect to COVID pandemic situation.

CONCLUSION

Mumbai’s Chhatrapati Shivaji Maharaj International Airport has solidified its position as India’s premier international aviation hub. The study's findings indicate that several key factors have contributed to this transformation:

- **Strategic Route Expansions:** New connections to Africa and Southeast Asia have diversified Mumbai's global reach.

- **Infrastructure Investments:** Enhanced cargo facilities and terminal upgrades have improved operational efficiency and capacity.

- **Resilience:** The rapid recovery post-COVID-19, outpacing global averages, demonstrates the airport's adaptability and the strong underlying demand for travel to and from Mumbai.

FUTURE SCOPE

Building upon the insights gained from this study, several avenues for future research can be explored:

- **Real-Time Analytics:** Integrate live flight data APIs for dynamic dashboards to provide up-to-the-minute insights into air traffic patterns.

- **Sustainability Metrics:** Analyze carbon emissions and green aviation initiatives at CSMIA to assess environmental impact and potential for sustainable growth.

- **Economic Impact Study:** Correlate passenger growth with tourism revenue and other economic indicators to quantify the airport's contribution to the regional and national economy.

REFERENCES

- Mumbai Aviation Traffic Data: OpenCity Mumbai.
- DGCA Annual Reports: DGCA Portal.
- Python Libraries: Pandas, Matplotlib, Seaborn.