A

Project on

**Indian Airlines**

Submitted in partial fulfillment of completion of the course

Advanced Diploma in IT, Networking and Cloud

Submitted by:

**Vinita Rawat**

**Sandhya Prajapati**

Under Guidance of:

**(Edunet Mentor) Mrs. Deepika**

**(Edunet Mentor) Miss. Ankita Shukla**

|  |  |  |
| --- | --- | --- |
| IBM-Logo - Chicago Innovation | DGT introduces high end diploma courses - digitalLEARNING Magazine | Edunet Foundation-Delhi- CSR Organization profile |

Year 2023

# Abstract

This project aims to conduct an in-depth analysis of data related to Indian airlines, with a focus on uncovering trends, patterns, and insights within the industry. The dataset encompasses various facets of airline operations, including but not limited to flight schedules, passenger demographics, pricing strategies, and operational performance.

The analysis will employ advanced statistical and machine learning techniques to extract meaningful information from the data. Key objectives include identifying factors influencing ticket prices, understanding the seasonal variations in travel demand, and evaluating the on-time performance of different airlines. Additionally, the project will delve into customer preferences, examining factors that contribute to passenger satisfaction and loyalty.

By leveraging this data-driven approach, the study aims to provide valuable insights for both industry stakeholders and policymakers. The findings are expected to contribute to strategic decision-making processes within the Indian aviation sector, ultimately fostering improvements in operational efficiency, customer experience, and overall industry competitiveness.

Top of Form

# Acknowledgement

I would like to express my heartfelt gratitude to all those who have contributed to the successful completion of this data analysis project on Indian airlines.

First and foremost, my project supervisor, for their guidance, support, and valuable insights throughout the entire research process. Their expertise and encouragement have been instrumental in shaping the direction of this project.

I am grateful to the data providers and organizations that made this research possible by sharing their valuable datasets and resources. Their cooperation has been crucial in ensuring the accuracy and relevance of the analysis.

I would also like to acknowledge the contributions of my colleagues and peers who provided constructive feedback and engaged in insightful discussions, enhancing the quality of the research.

Last but not least, I want to express my appreciation to my friends and family for their unwavering support and understanding during the course of this project. Their encouragement has been a source of motivation, and I am truly grateful for their presence in my life.

This research would not have been possible without the collective efforts and support of these individuals and organizations. Thank you all for being an integral part of this journey.

Top of Form

.

# Team Composition and Workload Division

1. Vinita – UI (front-end) & Database (backend)
2. Sandhya – UI (front-end) & Database (backend)

# Introduction to Problem

The aviation industry stands at the crossroads of innovation and complexity, with data playing a pivotal role in shaping its trajectory. In this era of unprecedented technological advancements, understanding the intricate dynamics of the Indian aviation sector becomes paramount. This project embarks on a comprehensive data analysis journey, seeking to unveil the hidden patterns and trends within the realm of Indian airlines.

India, with its diverse geography and burgeoning population, presents a unique landscape for air travel. The dataset under scrutiny encapsulates a myriad of dimensions, ranging from flight schedules and pricing strategies to passenger demographics and operational performance. As we delve into this rich tapestry of information, our primary goal is to extract actionable insights that can inform strategic decisions, enhance operational efficiency, and elevate the overall passenger experience.

The scope of the analysis spans the vast skies above, aiming to answer crucial questions that resonate with both industry stakeholders and the flying public. What factors influence ticket prices, and how do they evolve over time? Are there discernible patterns in travel demand, and how do different airlines navigate seasonal variations? Beyond the numbers, we aim to understand the pulse of passenger satisfaction—identifying the factors that contribute to a seamless and enjoyable travel experience.

Through the lens of advanced statistical and machine learning techniques, we aspire to not only unravel the complexities of the Indian aviation landscape but also contribute to the ongoing dialogue surrounding its future. This project serves as a beacon, illuminating the path towards informed decision-making, operational excellence, and a sky where every journey is as remarkable as the destination itself.

# Proposed Solution

As we embark on the ambitious journey of analyzing the intricacies of the Indian aviation sector, a robust and strategic approach is imperative to derive meaningful insights. Our proposed solution involves a multi-faceted methodology, leveraging advanced analytical tools and techniques to unravel the patterns embedded in the dataset. Here's a breakdown of our proposed solution:

1. Data Preprocessing and Cleaning:
   * Conduct a thorough examination of the dataset to identify and rectify any missing or inconsistent data.
   * Standardize and clean variables to ensure uniformity and accuracy in subsequent analyses.
2. Exploratory Data Analysis (EDA):
   * Employ descriptive statistics and visualization techniques to gain an initial understanding of the dataset's characteristics.
   * Identify outliers, anomalies, and trends that may influence the overall analysis.
3. Feature Engineering:
   * Extract relevant features that can provide deeper insights into the factors influencing ticket prices, seasonal variations, and operational performance.
   * Create new variables that capture the essence of passenger preferences and satisfaction.
4. Statistical Analysis:
   * Utilize statistical methods to assess the significance of identified trends and patterns.
   * Conduct hypothesis testing to validate assumptions and draw meaningful conclusions.
5. Machine Learning Models:
   * Develop predictive models to forecast ticket prices based on historical data and external factors.
   * Explore clustering algorithms to categorize airlines based on performance and customer satisfaction metrics.
6. Time Series Analysis:
   * Investigate temporal patterns in flight schedules and passenger demand.
   * Implement time series forecasting techniques to predict future trends and potential challenges.
7. Interactive Dashboards:
   * Create user-friendly dashboards for stakeholders to interactively explore and visualize the results.
   * Facilitate dynamic exploration of data insights, empowering decision-makers with real-time information.
8. Recommendation Engine:
   * Develop a recommendation engine to suggest strategies for optimizing pricing, improving operational efficiency, and enhancing customer satisfaction.

By seamlessly integrating these components, our proposed solution aims to not only dissect the complexities of the Indian airline industry but also provide actionable insights that can drive informed decision-making and contribute to the continuous evolution of this dynamic sector.

Top of Form

# Requirements

## Technology Stack

### Full stack development

Full stack development refers to the development of both front end (client side) and back end (server side) portions of web application. We have used Jupiter Notebook (with python & pandas library Kaggle as the database.

## Hardware

* + - Desktop/Laptop
    - Minimum 4GB RAM
    - Processor 64-bit
    - Hard Drive 250GB
    - Internet Connection

## Software

### Jupiter Notebook

### Jupiter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. It is widely used in data science, machine learning, scientific research, and education. The name "Jupiter" is a combination of three core programming languages it supports: Julia, Python, and R.

### Front End

HTML: Hyper Text Markup Language (HTML) is a markup language for creating a webpage. Webpages are usually viewed in a web browser. They can include writing, links, pictures, and even sound and video. HTML is used to mark and describe each of these kinds of content so the web browser can display them correctly.

CSS: CSS is the language for describing the presentation of Web pages, including colors, layout, and fonts. It allows one to adapt the presentation to different types of devices, such as large screens, small screens, or printers. CSS is independent of HTML and can be used with any XML-based markup language.

# User Requirements

* + - Electronic Device: Mobile, Laptop, Desktop or Tablet
    - Email Account
    - Access to Internet

1. **Implementation Details**

|  |
| --- |
| Importing libraries A library is a collection of functions that can be added to your Python code and called as necessary, just like any other function. Loading dataset Data loading is the process of copying and loading data or data sets from a source file, folder or application to a database or similar application. It is usually implemented by copying digital data from a source and pasting or loading the data to a data storage or processing utility. Viewing and understanding the Data Data understanding is the knowledge that you have about the data, the needs that the data will satisfy, its content and location. To be clear, it is much more than current location and a definition of what a data element means in situ within an application or data base. Data Visualization Data visualization is the representation of data through use of common graphics, such as charts, plots, info graphics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand. |

# 6. Future Scope

The culmination of this data analysis project not only marks the attainment of valuable insights into the present state of the Indian aviation industry but also lays the groundwork for future exploration and enhancement. The following avenues represent potential future scopes for extending and refining the impact of our analysis:

Dynamic Predictive Modeling:

Implement dynamic predictive models that can adapt to evolving market conditions and unforeseen external factors.

Explore machine learning algorithms that continuously learn from real-time data, enabling airlines to proactively adjust strategies.

Integration of External Data Sources:

Incorporate additional external data sources, such as economic indicators, weather patterns, and geopolitical events, to enhance the accuracy and scope of predictions.

Collaborate with relevant industries to access data that can contribute to a more holistic understanding of the aviation ecosystem.

Customer Sentiment Analysis:

Integrate natural language processing (NLP) techniques to analyze customer reviews, feedback, and sentiments.

Develop sentiment analysis models to gauge public perception and sentiment towards different airlines, aiding in reputation management.

Sustainability Metrics:

Integrate sustainability metrics into the analysis to evaluate the environmental impact of different airlines.

Explore ways to optimize operational efficiency and reduce carbon footprints, aligning with the growing focus on eco-friendly aviation practices.

Collaboration with Regulatory Bodies:

Collaborate with aviation regulatory bodies to incorporate regulatory changes and compliance requirements into the analysis.

Provide insights that can inform policy decisions and regulatory frameworks, contributing to the sustainable growth of the aviation industry.

Personalized Travel Experience:

Develop personalized recommendation systems for passengers, offering customized travel experiences based on historical preferences and behavior.

Enhance customer loyalty programs by tailoring rewards and incentives to individual passenger profiles.

Global Comparative Analysis:

Extend the analysis to include a comparative study with international aviation trends.

Identify best practices and strategies employed by global airlines, fostering a cross-cultural exchange of knowledge and innovation.

Real-time Analytics Dashboard:

Evolve the interactive dashboards into real-time analytics platforms, providing stakeholders with up-to-the-minute insights for agile decision-making.

As we envision the future scope of this project, the aim is not only to adapt to the changing landscape of the aviation industry but also to actively shape its trajectory. By embracing emerging technologies and forging collaborations, we can contribute to an aviation ecosystem that is resilient, sustainable, and responsive to the evolving needs of both airlines and passengers.

# 7. Conclusion

In conclusion, our deep dive into the data of Indian airlines has unveiled a roadmap for transformative growth. Rooted in advanced analytics and machine learning, our solutions pave the way for adaptive decision-making and strategic planning. As we conclude this phase, the horizon widens with opportunities for dynamic predictive modeling, sustainable practices, and global collaborations. The journey doesn't end; it transforms into a commitment to crafting an aviation narrative that blends efficiency, sustainability, and passenger-centricity. The skies of possibility beckon, and our insights serve as a compass, guiding the industry towards a future where every flight is an experience, and data propels progress. The journey continues, inviting us to soar higher and explore the boundless horizons of an evolving aviation landscape.

Top of FormIn conclusion, our deep dive into the data of Indian airlines has unveiled a roadmap for transformative growth. Rooted in advanced analytics and machine learning, our solutions pave the way for adaptive decision-making and strategic planning. As we conclude this phase, the horizon widens with opportunities for dynamic predictive modeling, sustainable practices, and global collaborations. The journey doesn't end; it transforms into a commitment to crafting an aviation narrative that blends efficiency, sustainability, and passenger-centricity. The skies of possibility beckon, and our insights serve as a compass, guiding the industry towards a future where every flight is an experience, and data propels progress. The journey continues, inviting us to soar higher and explore the boundless horizons of an evolving aviation landscape.

Top of Form

# Appendix A

## Project Code

# Indian Airlines Ticket Price Analysis

\* Defining the problem statement

\* Collecting the data

\* Exploratory data analysis

### 1) Defining the problem statment

In this project, we study the data which is in tabular format using various Python libraries like Pandas, Numpy, Matplotlib and Seaborn.

We study different columns of the table and try to co-relate them with others and find a relation between those two.

We try to find and analyze those key factors like class of travel, duration of fight, etc. which helps us understand the pricing of tickets to plan and schedule our air travel in efficient way

### 2) Collecting the data

##### Import the required Python libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

##### Reading the dataset using Pandas

df = pd.read\_csv("Indian Airlines.csv")

df.head(10)

Airline: Categorical, 6 airlines.

Flight: Categorical flight code.

Source City: Categorical, 6 cities.

Departure Time: Categorical time bins (6 labels).

Stops: Categorical, 3 values.

Arrival Time: Categorical time bins (6 labels).

Destination City: Categorical, 6 cities.

Class: Categorical, Business/Economy.

Duration: Continuous, hours.

Days Left: Derived from trip and booking dates.

Price: Target variable.

df.nunique()

for col in df:

if df[col].dtype == 'object':

print(df[col].unique())

##### About the columns:

1) In airline column there are 6 unique airlines: SpiceJet, AirAsia, Vistara, GO\_FIRST, Indigo, Air\_India

2) In source\_city & destination\_city there are 6 unique cities: Delhi, Mumbai, Bangalore, Kolkata, Hyderabad, Chennai

3) In arrival & departure columns there are 6 different timings: Night, Morning, Early\_Morning, Afternoon, Evening, Late\_Night

4) In class column there are 2 different classes: Economy, Business

### 3) Exploratory Data Analysis

##### 1. What are number of flights operated by each airline?

from turtle import title

plt.figure(figsize=(15,5))

NF = sns.countplot(x='airline', data = df)

NF.set(xlabel='Airline in India', ylabel='No. of flights', title='No. of flight by Airlines')

plt.show(NF)

From the above figure, we can see 'Vistara' has maximum no. of fights followed by 'Air India' while 'Spice Jet' has least no. of flights

##### 2. What is price range according to class of travel?

from turtle import title

plt.figure(figsize=(15,5))

CE = sns.stripplot(x='price', y='class', data = df)

CE.set(xlabel='Ticket cost', ylabel='Class of Travel', title='Price range according to Class of Travel')

plt.show(CE)

From the above figure, we can see 'Economy' class tickets usually cost between 2500 - 22500 while 'Business' class tickets usually cost between 25000 - 95000

##### 3. What is availability of Tickets according to class of travel?

from turtle import title

plt.figure(figsize=(15,5))

TA = sns.countplot(x='class', data = df)

TA.set(xlabel='Class of Travel', title='Availability of Tickets according to Class of Travel')

plt.show(TA)

From the above figure, we can see that availabilty of 'Economy' tickets is almost twice than availibitly of 'Business' class tickets which is explained by the fact that only 2 airlines - 'Air India, Vistara' offer 'Business' class tickets while all airlines offer 'Economy' class tixkets.

##### 4. How do ticket prices vary across different airlines and class of travel?

plt.figure(figsize=(15,5))

AS = sns.barplot(x='airline', y='price', hue='class', data = df.sort\_values('price'))

AS.set(xlabel='Airlines in India', ylabel='Price of Ticket', title='Economy Vs Business Ticket Prices by Airlines')

plt.show(AS)

From the above figure, we can conclude that 'Air Asia' offers the cheapest 'Economy' class tickets while 'Indigo, 'Go First', 'Spice Jet' are almost similarly priced. Meanwhile 'Air India' and 'Vistara' are priced much higher than other 4 airlines which can be explained on the basis that 'Air India' and 'Vistara' are both FSCs while rest are LCCs.

'Business' class tickets for 'Vistara' cost much higher than 'Air India' which can be due to better service, quality of seats available on 'Vistara' as compared to 'Air India'

##### 6. How do airline ticket prices vary depending on when you buy them?

df\_temp = df.groupby(['days\_left'])['price'].mean().reset\_index()

plt.figure(figsize=(15,5))

ax = plt.axes()

sns.regplot(x = df\_temp.loc[df\_temp['days\_left'] == 1].days\_left, y = df\_temp.loc[df\_temp['days\_left'] == 1].price, data= df\_temp, fit\_reg= False, ax=ax)

sns.regplot(x = df\_temp.loc[(df\_temp['days\_left'] > 1) & (df\_temp['days\_left'] < 20)].days\_left, y = df\_temp.loc[(df\_temp['days\_left'] > 1) & (df\_temp['days\_left'] < 20)].price, data = df\_temp, fit\_reg= True, ax=ax)

sns.regplot(x = df\_temp.loc[df\_temp['days\_left'] >= 20].days\_left, y = df\_temp.loc[df\_temp['days\_left'] >= 20].price, data = df\_temp, fit\_reg= True, ax=ax)

ax.set(xlabel='Tickets booked before X days', ylabel='Price of Ticket', title='Airline ticket prices based on days left before buying the ticket')

plt.show(ax)

From the above figure, we can conclude that ticket price rise slowly till 20 days from the date of flight, then rise sharply till the last day, while dramatically reducing just 1 day before the date of flight. This can be explained by the fact that people usually buy flight tickets within 2-3 weeks of flight which generates more profits for airlines. On last day, prices show dramatic reduction as airlines hope to fill the flight completely due to increase the load factor and decrease the operational cost per passenger.

##### 7. How does price of ticket vary depending on duration of flight?

df\_temp2 = df.groupby(['duration'])['price'].mean().reset\_index()

plt.figure(figsize=(15,5))

PD = sns.scatterplot(x='duration', y='price', data = df\_temp2)

PD = sns.regplot(x='duration', y='price', data = df\_temp2, order = 2)

PD.set(xlabel='Duration of flight', ylabel='Price of Ticket', title='Average price depending on duration of flight')

plt.show(PD)

From the above figure, we can see that the relationship is not linear but can be approximated by second degree curve. We can see linear growth in prices as duration of flight increases till 20 and then lowering again.

Some outliers may be affecting the curve.

##### 8. How does ticket price vary according to departure time and arrival time?

plt.figure(figsize=(15,5))

plt.subplot (1,2,1)

sns.boxplot(data = df, x = 'departure\_time', y = 'price', showfliers = False).set(xlabel = 'Departure Time', ylabel = 'Price of Ticket', title ='Price of Ticket depending on time of departure')

plt.subplot (1,2,2)

sns.boxplot(data = df, x = 'arrival\_time', y = 'price', showfliers = False).set(xlabel = 'Arrival Time', ylabel = 'Price of Ticket', title = 'Price of Ticket depending on time of arrival')

plt.show()

From the above figure, we can conclude that flights departing late at night are cheapest while those arriving early morning and late night are cheap too. Flights departung in afternoon are relatively cheap as well.

##### 9. How does ticket price vary depending on source and destination?

ax = sns.relplot(x = 'destination\_city', y = 'price', col='source\_city', col\_wrap= 3, kind= 'line', data = df)

ax.fig.subplots\_adjust(top=0.9)

ax.fig.suptitle('Airline ticket prices based on the source and destination cities')

plt.show(ax)

From the above figure, we can conclude that flight departing from Delhi are usually cheaper which can be explained by the fact that Delhi being capital has very strong connectivity with every other city and more no. of freqencies resulting in cheaper ticket prices. Chennai-Bangalore seems to be the most expensive route to fly while Hyderabad is most expensive city to fly.

##### 10. How does price of tickets vary based on no. of stops and airline?

fig, axs = plt.subplots(1,2, gridspec\_kw= {'width\_ratios': [3,1]}, figsize = (15,5))

sns.barplot(y = 'price', x = 'airline', hue = 'stops', data = df.loc[df['class'] == 'Economy'].sort\_values('price', ascending= False), ax = axs[0])

axs[0].set(xlabel='Airlines', ylabel='Price of Ticket', title='Price of Airline tickets based on No. of Stops in Economy Class')

sns.barplot(y='price', x='airline', hue='stops', data= df.loc[df['class'] == 'Business'].sort\_values('price', ascending= False), ax = axs[1])

axs[1].set(xlabel='Airlines', ylabel='Price of Ticket', title='Price of Airline tickets based on No. of Stops in Business Class')

plt.show(fig, axs)

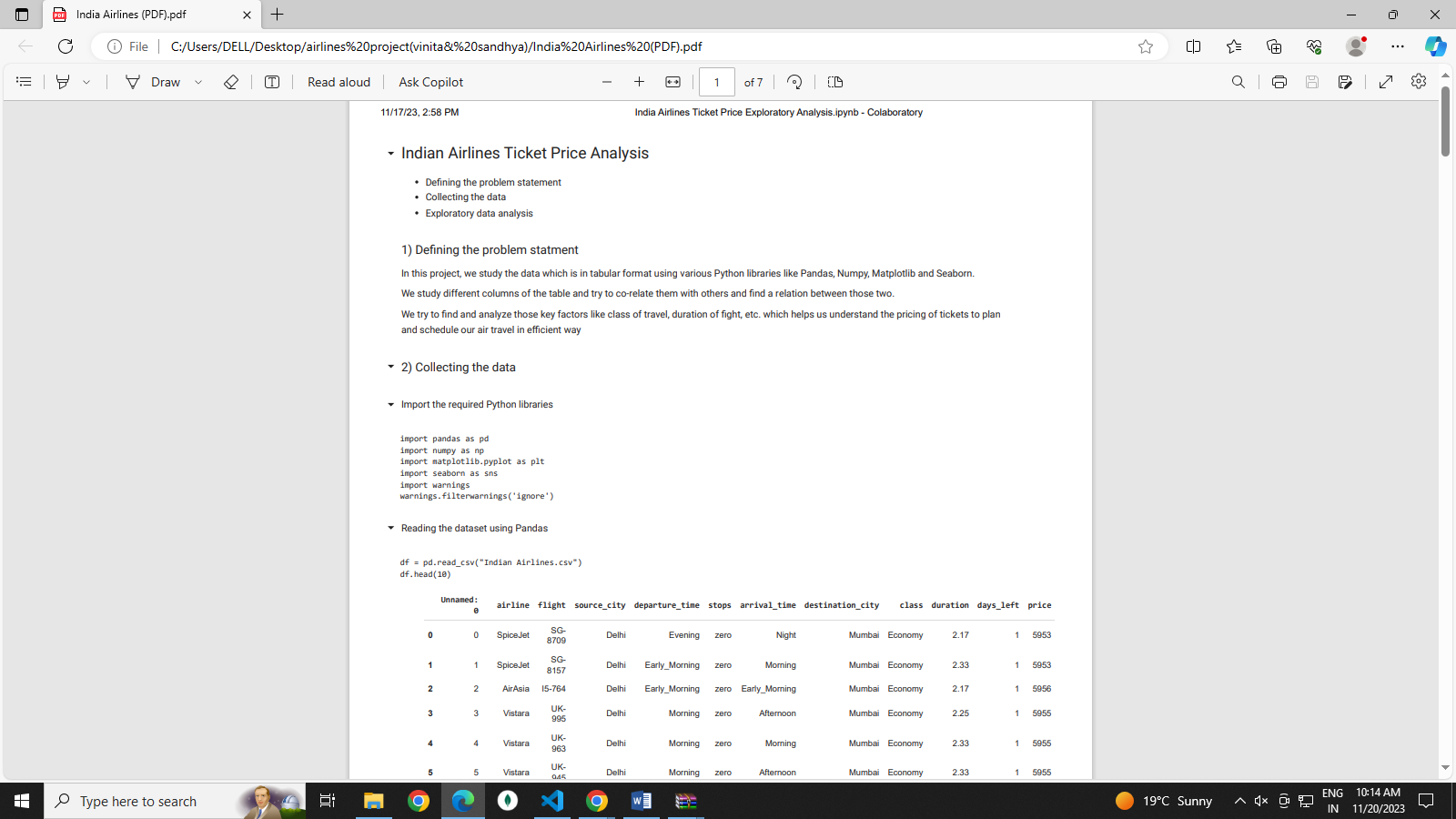
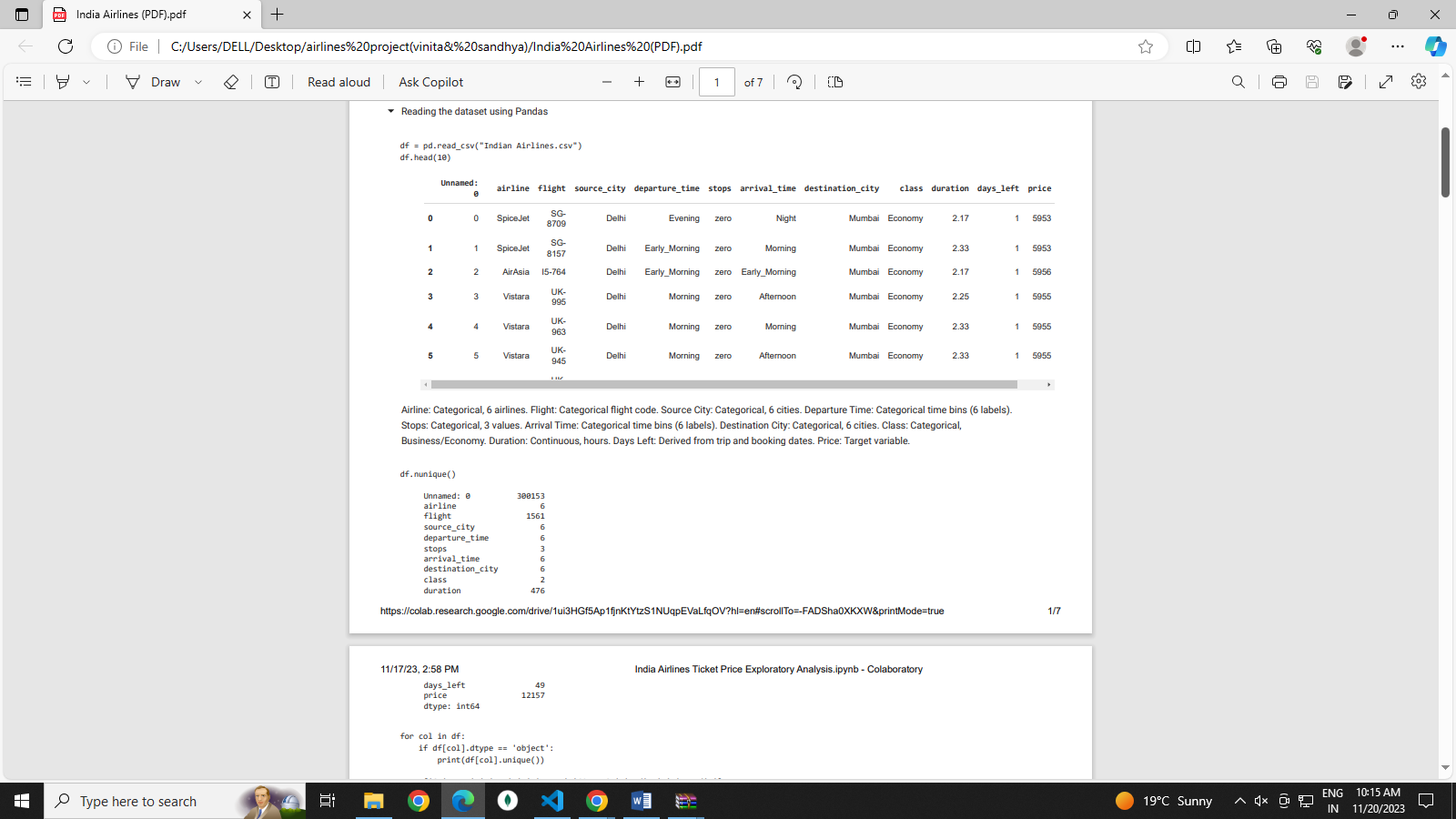
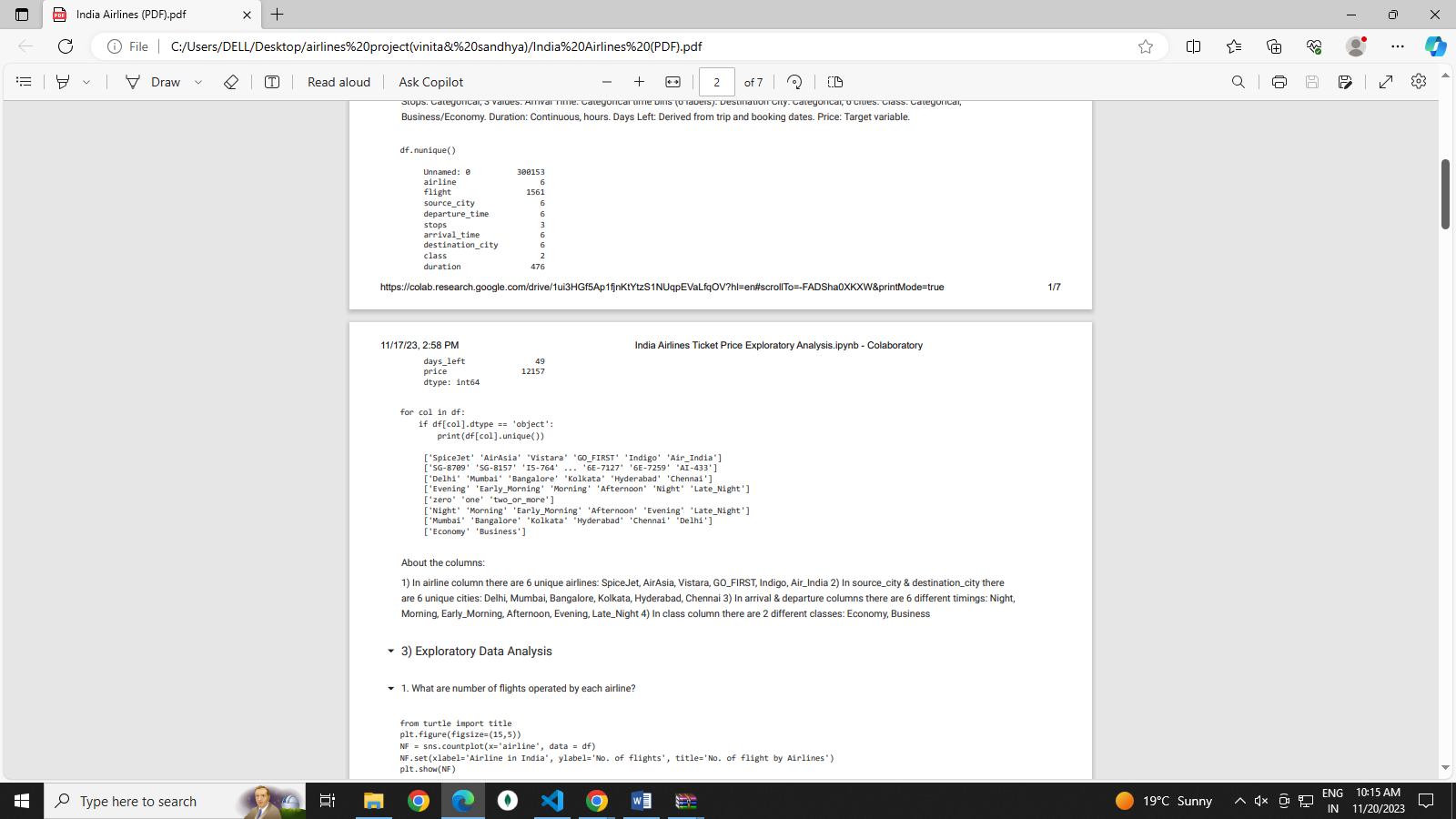
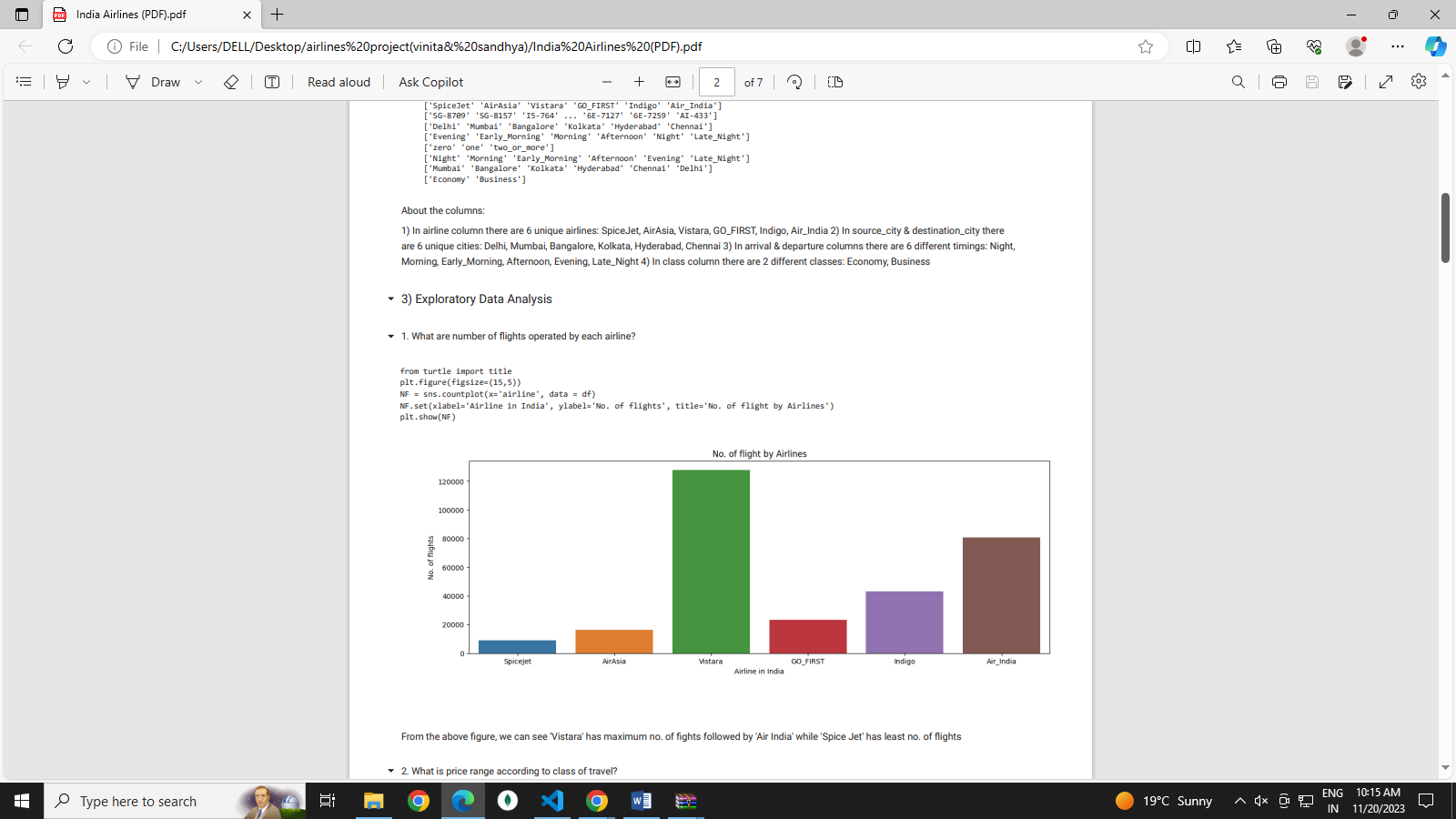
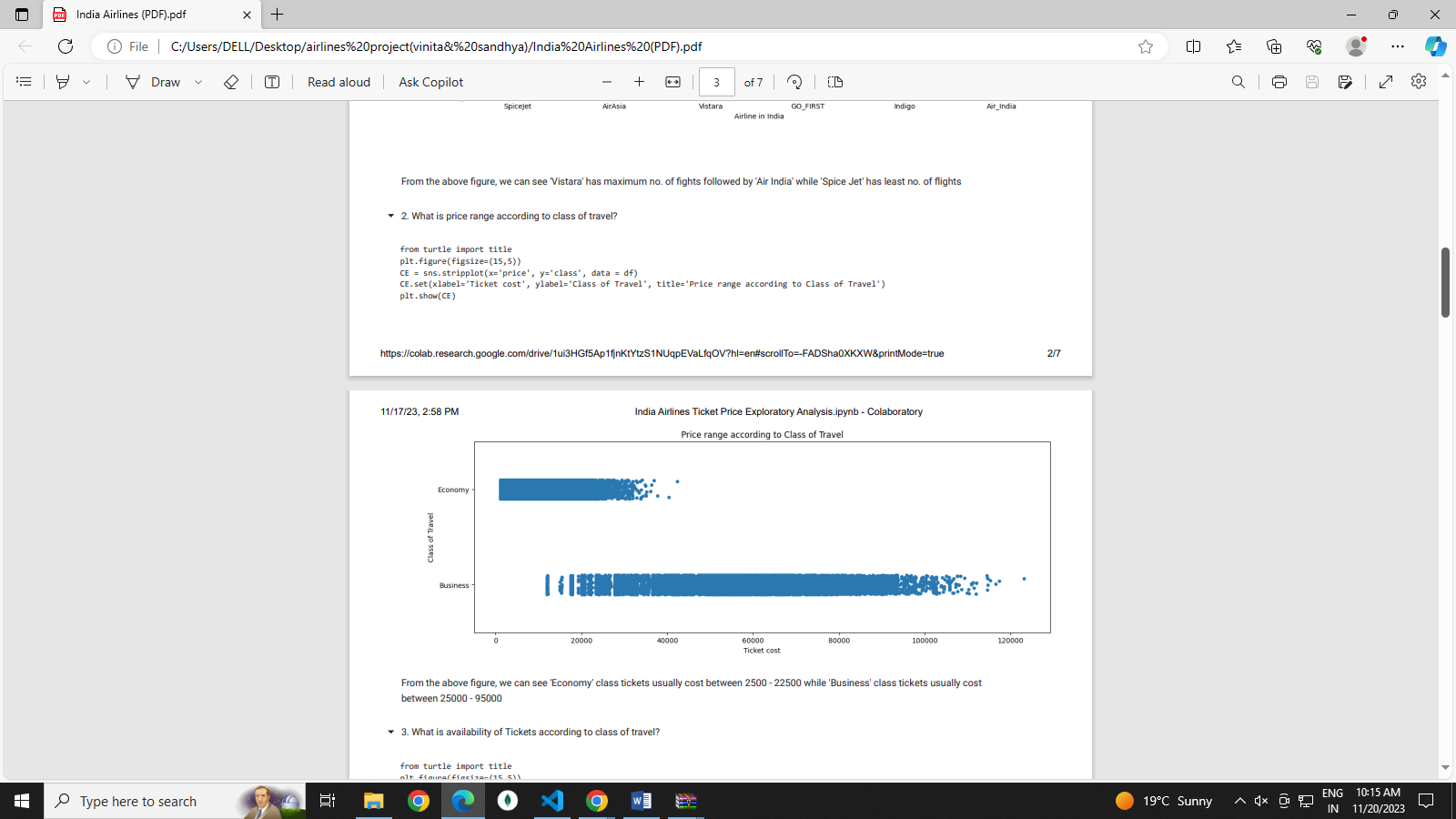
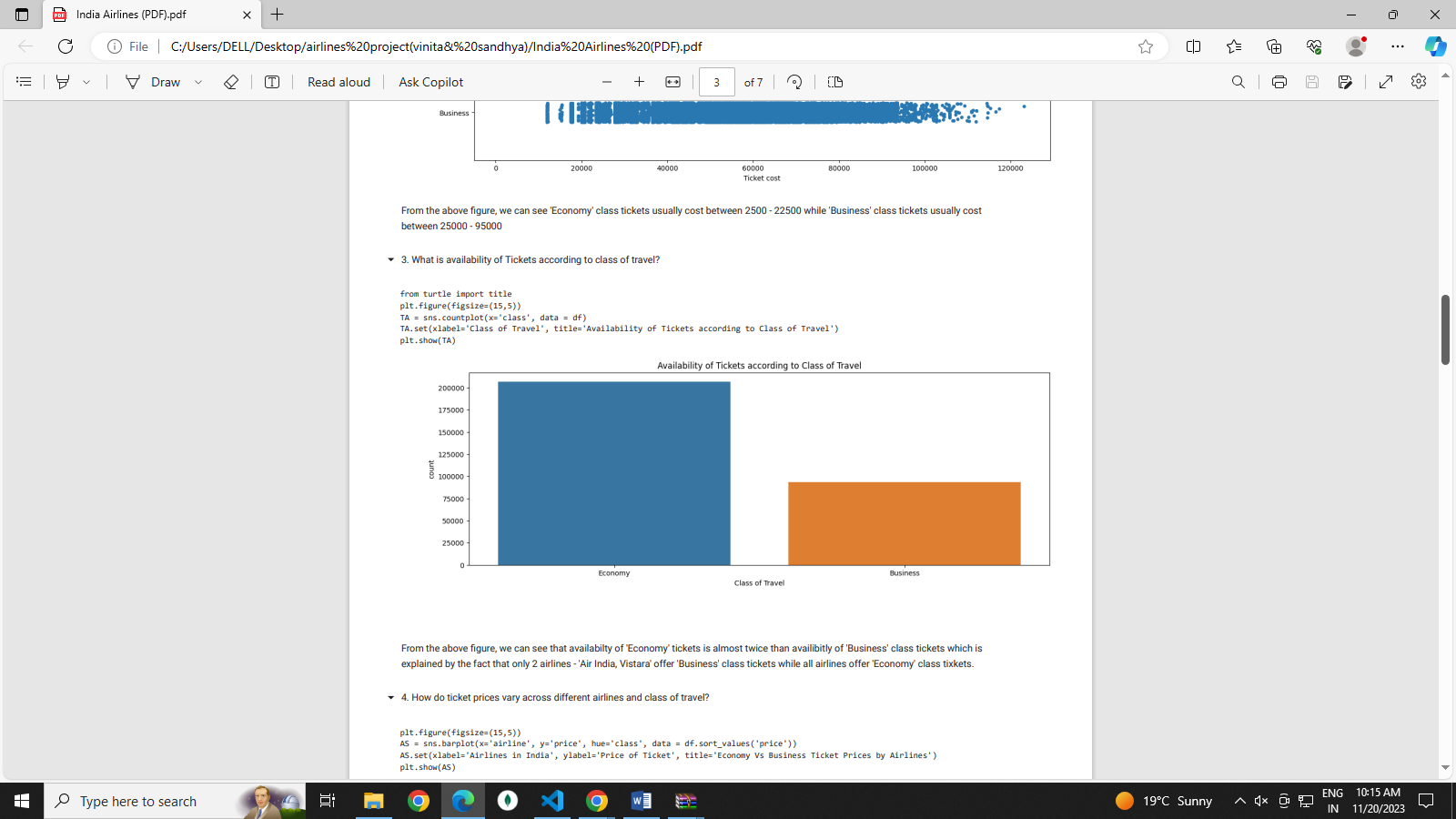
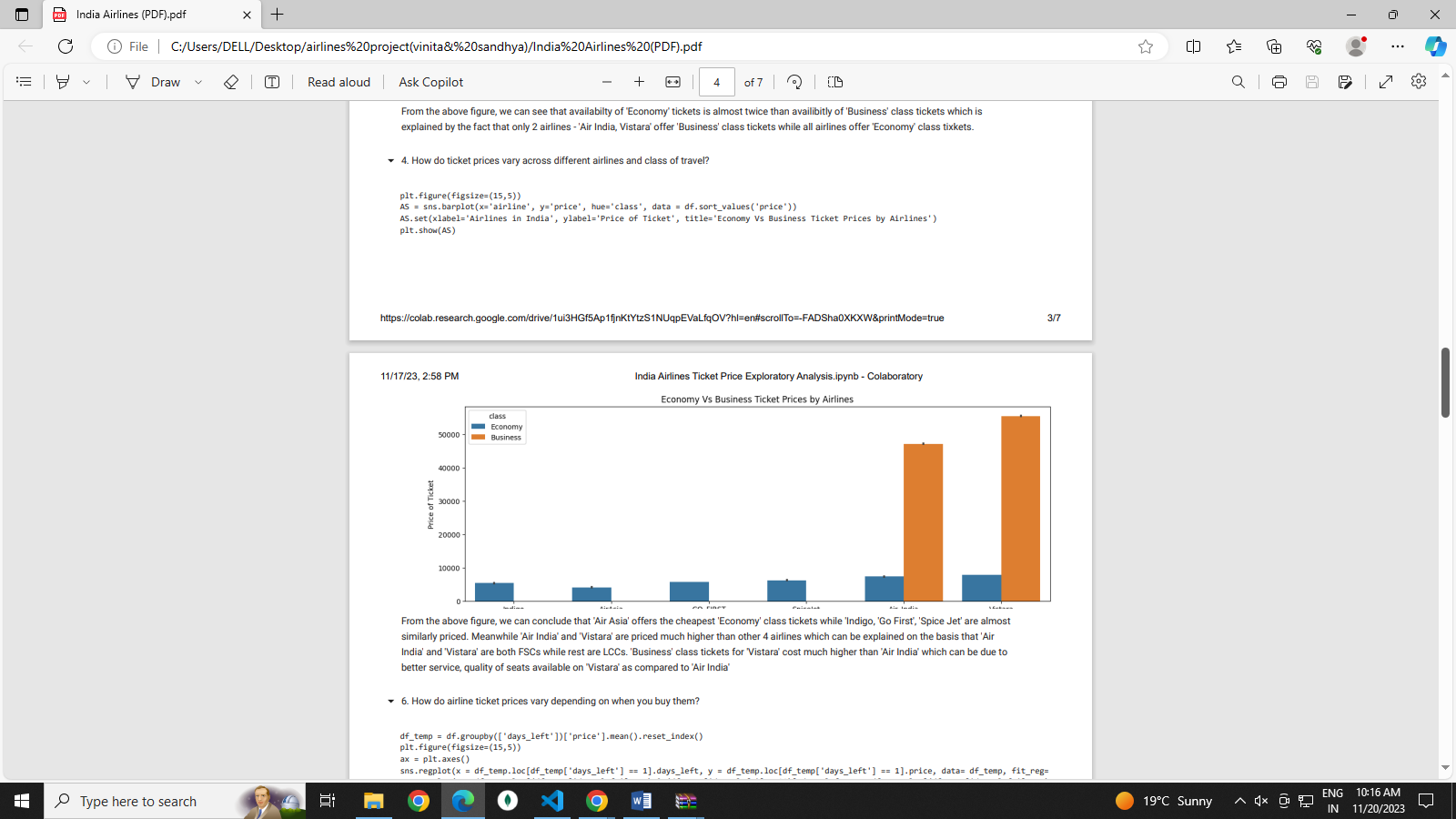
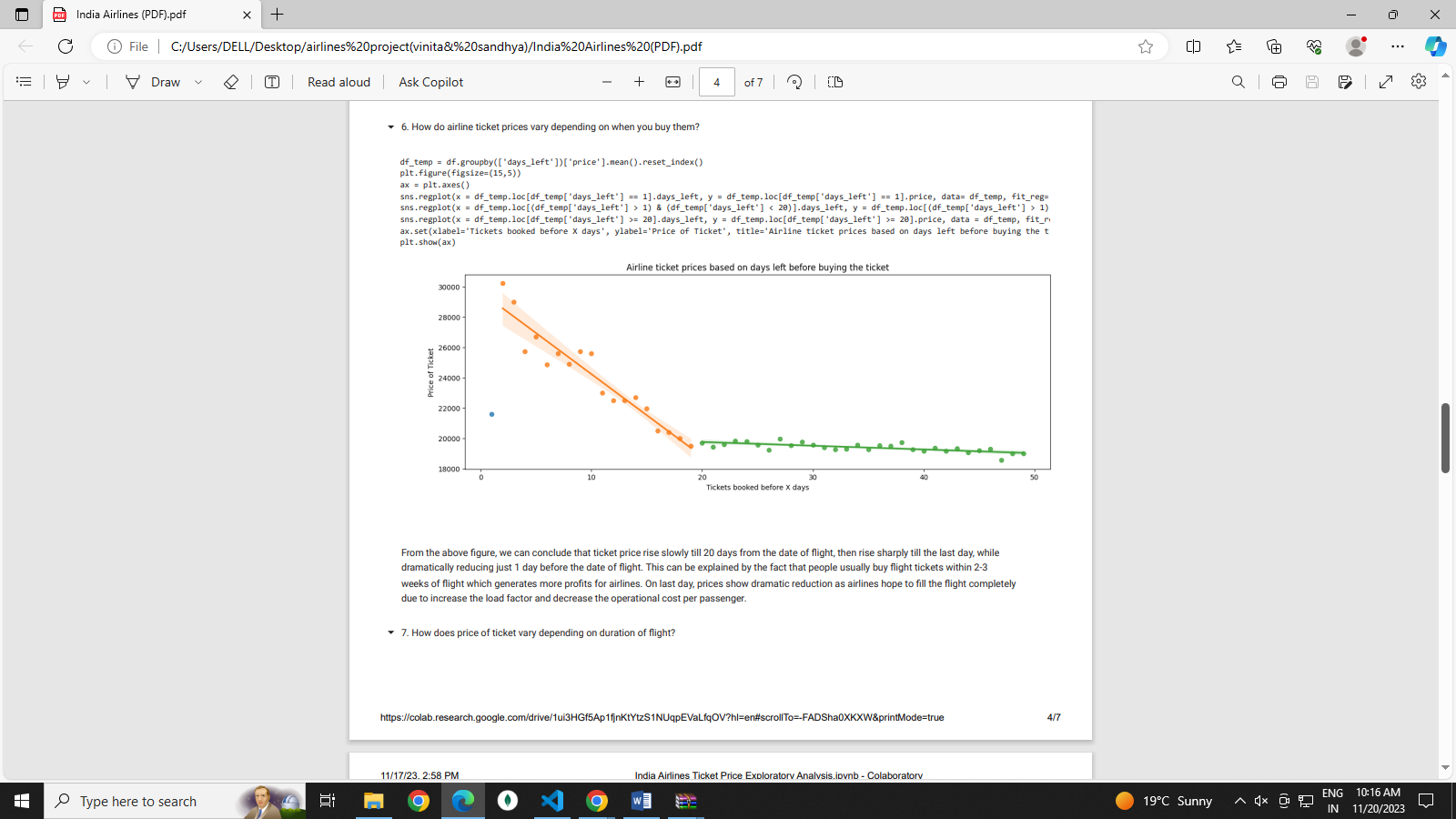
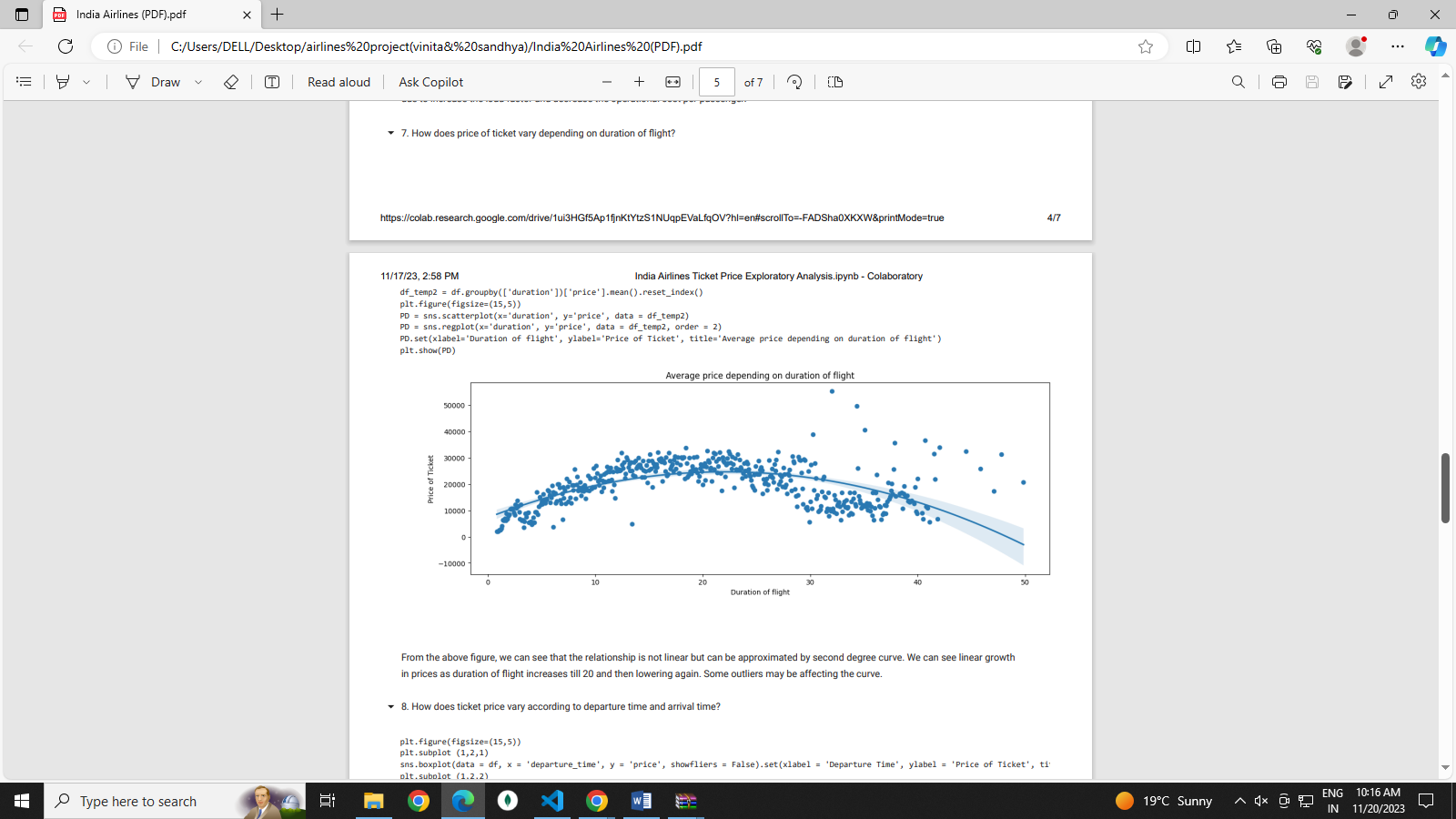
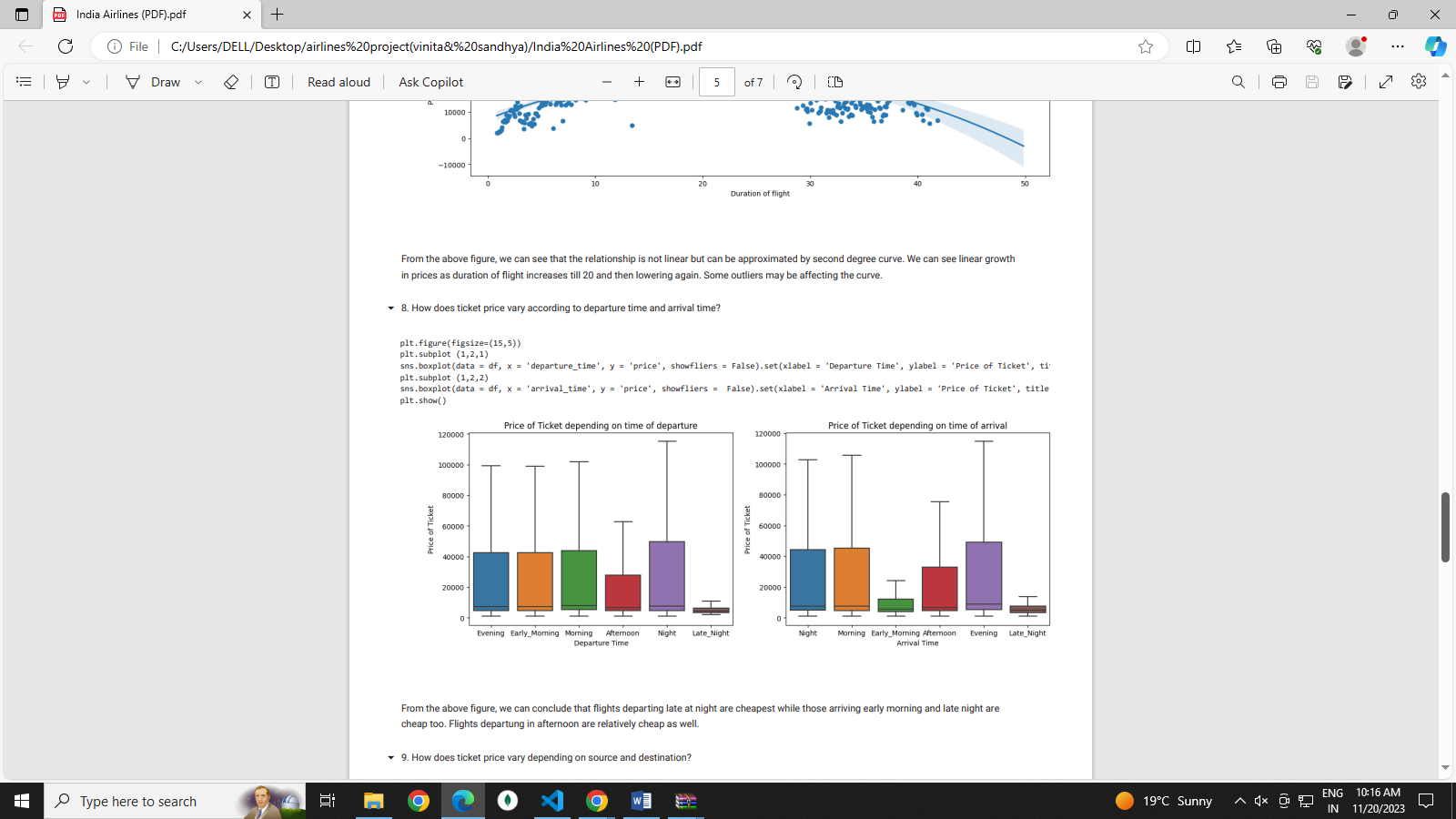
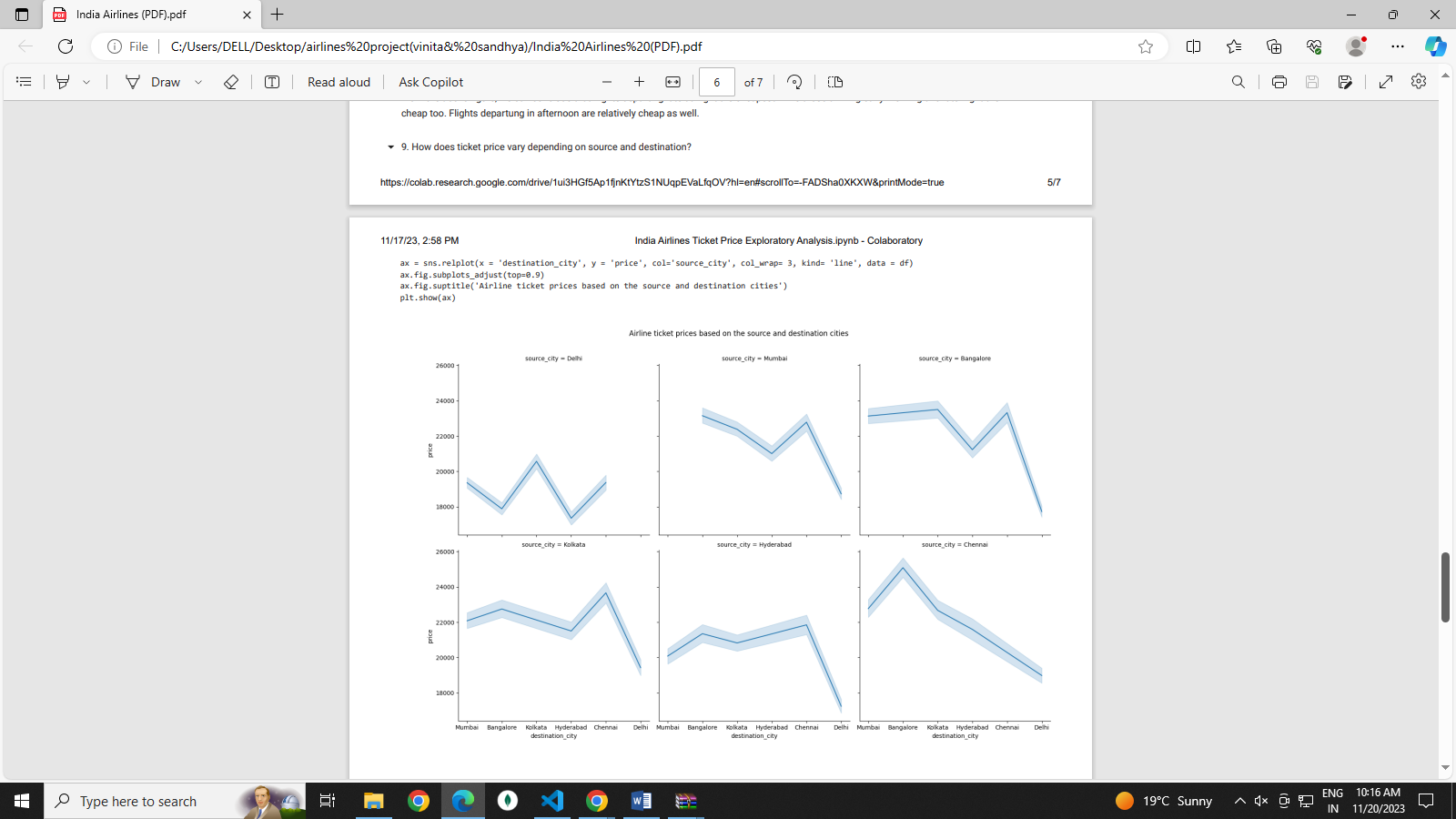
From the above figure, we can conclude that Non-Stop flights are generally the cheapest while One-Stop flights are more expensive and 2+ stop flights are most expensive which can be explained on basis that as one undertakes more flights to fly to destination, it costs more.

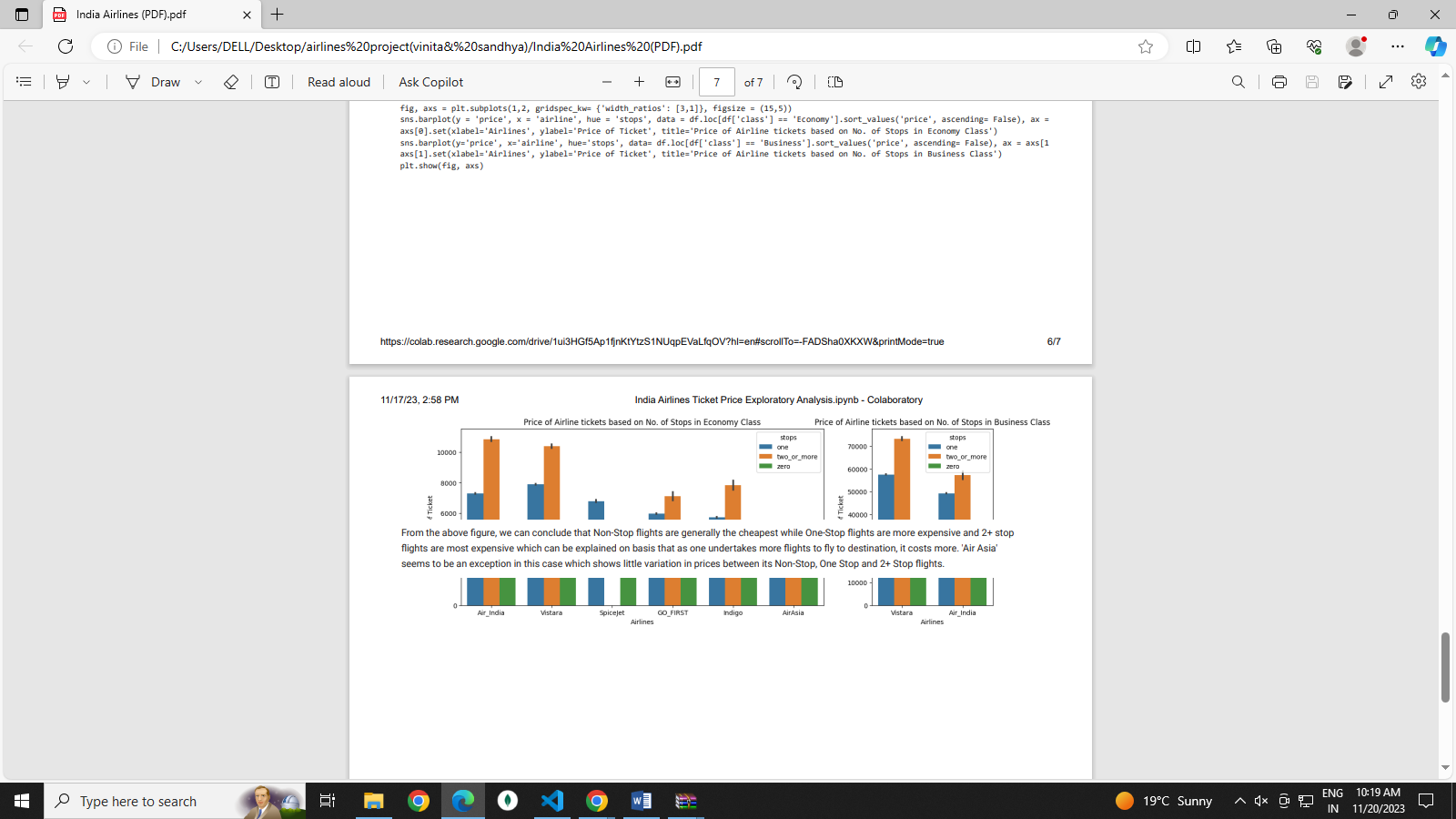
'Air Asia' seems to be an exception in this case which shows little variation in prices between its Non-Stop, One Stop and 2+ Stop flights.

# Appendix B

## Screenshot of Project

## Loading Dataset



**Thank you**