INDUSTRIAL INTERNSHIP REPORT

British Airways Data Science

Submitted in Partial fulfilment of the
requirements for the award of

Degree of **BTech Computer Science Engineering**

Submitted By:

Name: **Dhyey Bhatt**

University Roll No: **21BT04014**

Branch: **Computer Science and Engineering**

Submitted To:

Ms. Swati Saxena

School of Technology

GSFC University, Vadodara

Internship Institution: **Forage Platform**

Internship Period: **1 Month**

Date of Report Submission: **29-01-2024**

DECLARATION

I hereby declare that the Industrial Internship Report entitled “British Airways Data Science” is an authentic record of my own work as requirements of Industrial Internship during the period from 18-12-2023 to 13-01-2024 for the award of degree BTech Computer Science Engineering, GSFC University, Vadodara, under the guidance of Faculty Mentor Ms. Swati Saxena.

(Name of Student) Bhatt Dhyey NileshBhai
(University Roll No.) 21BT04014

Date: 29-01-2024

CERTIFICATE

This is to certify that Mr. Dhyey Bhatt has completed Industrial Internship during the period from 18/12/23 to 13/01/2024 in British Airways as a Partial Fulfilment of Internship. He was trained in the field of Data Science.

Signature & Seal of Host Company Mentor

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Ms. Swati Saxena, my faculty mentor at GSFC University in the field of Computer Science. Her unwavering support, insightful guidance, and dedication have played a pivotal role in shaping my academic journey. Ms. Saxena's wealth of knowledge, passion for the subject, and commitment to excellence have inspired and motivated me to strive for the highest standards in my studies.

I am truly grateful for the mentorship provided by Ms. Swati Saxena, as it has been instrumental in my personal and academic growth. I feel fortunate to have had such a dedicated mentor who has not only imparted knowledge but also instilled a passion for learning and innovation within me.

TABLE OF CONTENT

List of Figures

Abbreviations and Nomenclature

EDA	Exploratory Data Analysis
NLTK	Natural Language Tool-kit
Sklearn	Scikit-Learn
BA	British Airways
NLP	Natural Language Processing

Chapter: 1 Introduction

Embarking on an internship at British Airways Data Science was a transformative journey that seamlessly blended my passion for data science with the intricacies of the aviation industry. This opportunity afforded me a ringside view of the intersection between cutting-edge technology and the complex operational dynamics of one of the world's leading airlines.

Nestled within the heart of British Airways' analytical hub, my internship was marked by exposure to an environment teeming with innovation and collaborative spirit. From day one, I found myself working alongside seasoned professionals who seamlessly integrated theoretical knowledge with practical applications. The overarching goal was clear – harnessing the power of data to optimize every facet of airline operations.

My primary focus entered on unravelling the mysteries embedded in massive datasets. Flight patterns, passenger preferences, and operational nuances became the canvas on which I applied statistical models and machine learning algorithms. This hands-on experience not only solidified my technical skills but also provided a holistic understanding of the challenges inherent in the aviation landscape.

One of the most rewarding aspects of my internship was the direct contribution to projects aimed at enhancing overall performance and customer experience. Collaborating across cross-functional teams, I delved into the intricacies of flight scheduling, fuel efficiency, and predictive maintenance. The application of advanced analytics to solve real-world problems underscored the critical role data science plays in shaping the future trajectory of aviation.

Working with state-of-the-art technologies, I honed my abilities in data manipulation, visualization, and interpretation. The fast-paced nature of the aviation industry demanded agility, and I became adept at deriving actionable insights from data in real-time scenarios. The mentorship received from industry experts not only sharpened my technical acumen but also provided invaluable insights into the nuanced decision-making processes within the airline sector.

Beyond the technical realm, the internship at British Airways Data Science provided a unique vantage point into the corporate culture of a global aviation giant. I had the privilege of attending industry conferences, networking with professionals across departments, and gaining insights into the broader strategic vision of British Airways. This holistic exposure added a layer of understanding that extended beyond the confines of a traditional internship.

As I reflect on this transformative experience, I am not only equipped with a diverse skill set in data science but also possess a profound appreciation for the collaborative synergy required to navigate the complexities of the aviation industry. My time at British Airways Data Science has laid the foundation for a future career fuelled by a passion for data-driven innovation and a commitment to contributing meaningfully to the evolution of the aviation landscape.

Chapter: 2 Major components

There are two major components of Internship:

1. Scrape and analyse customer review data to uncover findings for British Airways
2. Build a predictive model to understand factors that influence buying behaviour

1. Scrape and Analyse Customer Review:

During my internship at British Airways, a significant task involved the extraction and analysis of customer review data to glean insights into the airline's performance and customer sentiments. The primary objective was to employ web scraping techniques to gather a comprehensive dataset of customer reviews related to British Airways. Subsequently, a meticulous analysis of this dataset was conducted, aiming to uncover valuable patterns, trends, and sentiments expressed by customers.

The initial phase of the task involved the utilization of web scraping tools to systematically collect a diverse range of customer reviews from various online platforms. This process required careful consideration of ethical guidelines and adherence to data privacy regulations to ensure the responsible collection of customer feedback.

Once the dataset was compiled, a robust analysis was undertaken to extract meaningful information. Techniques such as sentiment analysis, text mining, and natural language processing (NLP) were employed to decipher the sentiments and opinions embedded in the customer reviews. This allowed for the identification of key themes, common issues, and positive aspects that customers frequently highlighted.

The analysis not only focused on the sentiment polarity but also delved into the specifics of customer feedback. It involved categorizing comments into relevant topics, assessing the frequency of recurring issues, and identifying any emerging patterns. The goal was to provide British Airways with actionable insights that could inform strategic decision-making, service enhancements, and customer relationship management.

Furthermore, the analysis aimed to quantify the overall satisfaction levels of customers, allowing for a nuanced understanding of the factors influencing customer experiences. Through statistical methods and data visualization techniques, the findings were presented in a comprehensible format, enabling stakeholders to grasp the nuances of customer sentiments and make informed decisions.

This task was not only instrumental in providing a snapshot of the current customer landscape for British Airways but also laid the groundwork for potential improvements in service quality, customer satisfaction, and overall business strategy. The integration of web scraping and data analysis techniques demonstrated their applicability in gaining valuable business intelligence from unstructured data sources, illustrating the practical relevance of data science methodologies in the aviation industry.

2. **Build a Predictive Model:**

The task assigned during the internship at British Airways involved the development of a predictive model aimed at comprehending the various factors influencing buying behaviour. This multifaceted project required a strategic and analytical approach to uncover patterns and dependencies within the dataset.

The process initiated with a meticulous exploration of the available data, encompassing diverse variables related to customer interactions, preferences, and purchasing behaviours. The objective was to discern inherent relationships and correlations among these variables, ultimately identifying the key determinants that significantly impact customers' decision-making processes.

In the construction of the predictive model, a combination of statistical and machine learning methodologies was employed. This intricate blend allowed for a nuanced analysis of the dataset, leveraging both traditional statistical techniques and advanced machine learning algorithms to capture complex patterns and trends. The model was designed not only to predict buying behaviour but also to elucidate the underlying factors contributing to such behaviour.

The integration of relevant features and variables into the model aimed at capturing the essence of customer decision-making. This involved an in-depth exploration of demographic data, historical purchase patterns, customer feedback, and any other pertinent information that could provide valuable insights into the intricacies of buying behaviour within the aviation industry.

Validation and refinement of the predictive model were crucial stages in ensuring its accuracy and reliability. Rigorous testing methodologies were employed to assess the model's performance, allowing for adjustments and improvements based on feedback and real-world data outcomes. The iterative nature of this process underscored the commitment to developing a robust and effective tool for understanding and predicting buying behaviour.

The ultimate goal of this predictive model was to equip British Airways with actionable insights that could inform targeted marketing strategies, enhance customer experiences, and optimize business operations. By unravelling the nuanced interplay of factors influencing buying decisions, the model sought to contribute to the airline's ability to adapt and thrive in a competitive marketplace, ultimately fostering a more customer-centric and data-driven approach to decision-making within the organization.

Chapter: 3 Methodology

3.1. Project Overview:

- The methodology implemented during my internship at British Airways Data Science was designed to systematically extract nuanced insights from customer reviews and develop a robust predictive model to unravel the intricacies of buying behaviour. This comprehensive approach aimed to align with the overarching objectives of enhancing operational efficiency and customer satisfaction within the airline industry.

3.2. Research Design:

- The research design adopted for the project involved a careful blend of qualitative and quantitative methods. This approach was chosen to ensure a holistic understanding of customer sentiments through qualitative analysis of reviews and to derive quantitative insights through the development of a predictive model.

3.3. Data Collection:

- **Task-1 (Scraping and Analysing Customer Review Data):**
  - Employing web scraping techniques to collect diverse customer reviews from various online platforms.
  - Utilizing Python, with the strategic use of libraries such as Beautiful Soup and Scrapy, for efficient and targeted data extraction.
  - Conducting data pre-processing to ensure cleanliness, including text cleaning, duplicate removal, and dataset structuring.

- **Task-2 (Building a Predictive Model):**
  - Curating the dataset for the predictive model from internal records, capturing variables pertinent to buying behaviour.
  - Utilizing Python for data manipulation and pre-processing, with the aid of Pandas and NumPy.
  - Applying feature engineering techniques to distil meaningful insights from the dataset.

3.4. Analysis Methods:

- **Task-1:**
  - Employing sentiment analysis using Natural Language Processing (NLP) techniques, including tokenization and sentiment scoring.
  - Conducting geographical analysis through geospatial mapping and temporal trend analysis using time-series analysis.

- **Task-2:**
  - Utilizing machine learning techniques, with algorithms such as Random Forest and Gradient Boosting, for the development of the predictive model.
  - Conducting feature importance analysis to identify key factors influencing buying behavior.

3.5. Tools and Technologies:

- ➢ Leveraging Python as the primary programming language for both tasks.
- ➢ Utilizing libraries such as Scikit-learn, NLTK, and spaCy for natural language processing and machine learning.
- ➢ Employing Colab Notebooks for coding, analysis, and result visualization.
- ➢ Leveraging web scraping tools, including BeautifulSoup and Scrapy, for efficient extraction of customer reviews.

3.6. Ethical Considerations:

- ➢ Maintaining strict adherence to ethical guidelines throughout the project.
- ➢ Obtaining explicit consent for web scraping activities.
- ➢ Prioritizing privacy in the handling of customer reviews through anonymization processes.

3.7. Limitations:

- ➢ Acknowledging potential biases in online customer reviews.
- ➢ Recognizing the subjectivity inherent in sentiment analysis.
- ➢ Navigating constraints related to data availability and quality.

3.8. Project Timeline:

- ➢ Structuring the project into phases, including data collection, preprocessing, analysis, and model development.
- ➢ Integrating regular checkpoints and adjustments into the timeline for agility in response to unforeseen challenges.

In summation, this methodology embraced a holistic and multidimensional approach, seamlessly intertwining technological prowess, ethical considerations, and analytical depth. It provided British Airways Data Science with a rich tapestry of insights, aligning the diverse facets of data science to unravel the complexities within the aviation industry.

Chapter: 4 Tools and Technology used

1. **Python:**

Python, developed by Guido van Rossum in the late 1980s, has evolved into a versatile and widely-used programming language known for its simplicity, readability, and extensive community support. Emphasizing code readability, Python's clean syntax and use of indentation for code blocks contribute to a structured coding style. This simplicity, coupled with an extensive standard library, makes Python accessible to beginners and accelerates development for experienced programmers.

Python's community is vibrant and inclusive, evident in the Python Package Index (PyPI) hosting numerous third-party packages. The language's extensive documentation further supports developers of all levels. Its versatility is demonstrated across various domains, from web development (Django, Flask) and scientific computing (NumPy, SciPy) to machine learning (Scikit-learn, TensorFlow) and automation.

In data science, Python is dominant, with libraries like Pandas, NumPy, and SciPy offering robust data manipulation and scientific computing capabilities. For machine learning, frameworks like Scikit-learn, TensorFlow, and PyTorch leverage Python's simplicity for developing sophisticated models.

Python frameworks, such as Django and Flask, are popular in web development for their simplicity, scalability, and adherence to the Don't Repeat Yourself (DRY) principle. Django, a high-level web framework, follows the Model-View-Controller (MVC) pattern, streamlining the development of web applications.

Python's scripting capabilities make it a preferred choice for automation tasks and system administration. Its portability and libraries like os and shutil simplify file operations, while modules like subprocess facilitate system interaction.

Python's open-source philosophy and community-driven development model contribute to its continuous improvement and widespread adoption. This collaborative nature fosters innovation, leading to a rich ecosystem of tools and libraries.

In conclusion, Python's simplicity, versatility, and strong community make it a dominant force in programming. Whether applied in web development, data analysis, or machine learning, Python's adaptability and ease of use make it an indispensable tool for developers across diverse domains. As technology advances, Python's influence continues to grow, cementing its role as a language that empowers developers to transform ideas into reality.

2. **Google Colab:**

Google Colab, a cloud-based platform, has emerged as a powerful tool for Python-based development, particularly in data science and machine learning projects. Providing a collaborative and browser-based interface, Colab facilitates seamless integration with Google Drive, enabling easy sharing and collaboration on Jupyter notebooks.

With its GPU support, Google Colab accelerates computation-intensive tasks, making it an attractive choice for machine learning practitioners. The platform supports Python, and its integration with popular libraries such as NumPy, Pandas, and TensorFlow streamlines data analysis and model development workflows.

Colab's cloud-based nature eliminates the need for local hardware with significant computing power, making it accessible to a broader audience. Its collaborative features allow multiple users to work on the same notebook simultaneously, fostering teamwork in data science projects.

Developers appreciate the convenience of running code directly in the browser without worrying about local setup and dependencies. Colab also supports the installation of additional libraries, providing flexibility for diverse project requirements.

Furthermore, Colab offers pre-installed libraries for machine learning, including Scikit-learn and Matplotlib, reducing setup time for practitioners. The platform's integration with Google Sheets and BigQuery enhances its capabilities for data manipulation and analysis.

In summary, Google Colab serves as an efficient and collaborative environment for Python-based data science and machine learning projects. Its cloud-based infrastructure, GPU support, and integration with popular Python libraries make it a valuable asset for developers and researchers, streamlining the process of experimentation, collaboration, and sharing of code and insights.

3. **Scikit-learn:**

Scikit-learn, a powerful machine learning library for Python, plays a pivotal role in the development and implementation of machine learning models. As an integral part of the Python ecosystem, Scikit-learn simplifies the machine learning pipeline with its extensive set of tools for data preprocessing, model selection, and evaluation. Its user-friendly API enables quick experimentation with different algorithms and parameters, making it a preferred choice for both beginners and experienced data scientists.

Scikit-learn's versatility is evident in its support for various machine learning tasks, including classification, regression, clustering, and dimensionality reduction. The library provides a consistent interface for different algorithms, allowing users to seamlessly switch between models without significant code changes. This abstraction enhances code readability and facilitates the comparison of different algorithms for a given task.

The library's emphasis on ease of use extends to its straightforward integration with other popular Python libraries such as NumPy, SciPy, and Matplotlib. This integration streamlines the incorporation of Scikit-learn into existing data analysis and visualization workflows, enabling a holistic approach to machine learning projects.

Scikit-learn's comprehensive documentation and wealth of educational resources make it accessible to a broad audience. The library's commitment to best practices and model evaluation ensures that users can make informed decisions during the model development process. From hyperparameter tuning to cross-validation, Scikit-learn provides tools to enhance model performance and generalization to unseen data.

In the rapidly evolving field of machine learning, Scikit-learn remains a stalwart companion for data scientists and researchers. Its robust implementation of diverse algorithms, consistent API design, and commitment to simplicity contribute to its widespread adoption and continued relevance. As machine learning continues to shape various industries, Scikit-learn's role as a reliable and versatile library solidifies, empowering practitioners to harness the power of machine learning for diverse applications.

4. NLTK (Natural Language Toolkit):

NLTK, or the Natural Language Toolkit, is a Python library designed for natural language processing (NLP). Its development aimed to facilitate tasks such as text analysis, sentiment analysis, and language understanding. NLTK is an essential tool for researchers, educators, and developers working on projects that involve the manipulation and analysis of human language.

NLTK provides a wide range of modules for various NLP tasks, including tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition. The toolkit's modular design allows users to pick and choose the components relevant to their specific NLP needs, promoting flexibility and efficiency.

One of NLTK's strengths lies in its extensive collection of corpora and lexical resources. These resources, ranging from annotated text collections to lexical databases, empower users with diverse datasets for training and evaluation. The availability of such comprehensive linguistic resources enhances the accuracy and depth of NLP applications developed using NLTK.

The toolkit's user-friendly interface and well-documented API make it accessible to both beginners and experienced NLP practitioners. Its integration capabilities with other Python libraries, such as NumPy and Scikit-learn, contribute to a seamless workflow for comprehensive data analysis and machine learning applications involving natural language.

NLTK has played a crucial role in advancing research and applications in the field of NLP. From educational initiatives to industry applications, NLTK has become a go-to tool for projects involving the processing and understanding of human language. The community support around NLTK ensures that the toolkit remains updated and aligned with the latest advancements in NLP.

In conclusion, NLTK stands as a powerful and versatile toolkit for natural language processing in Python. Its rich feature set, extensive corpora, and user-friendly design make it an invaluable resource for anyone involved in linguistic research, text analysis, or the development of NLP applications. As the field of NLP continues to evolve, NLTK remains at the forefront, contributing to advancements in language technology and the broader applications of natural language understanding.

5. BeautifulSoup:

BeautifulSoup, a Python library for web scraping, provides an effective tool for extracting structured data from HTML and XML files. Its simplicity and flexibility make it a preferred choice for parsing and navigating through HTML documents.

Developed to simplify the extraction of information from web pages, BeautifulSoup aids in the creation of web scrapers by offering intuitive methods for searching, filtering, and navigating HTML or XML trees. Its usage does not require extensive knowledge of HTML, allowing developers to focus on extracting relevant data.

BeautifulSoup facilitates the extraction of data from websites by providing methods to locate and filter HTML elements based on tags, attributes, and text content. This enables the isolation of specific data points within a webpage, streamlining the process of collecting valuable information.

Additionally, the library's ability to handle poorly formatted HTML and gracefully navigate through complex document structures contributes to its reliability. It effectively transforms unstructured web data into a format suitable for further analysis and processing within a Python environment.

In conjunction with Python's requests library, BeautifulSoup becomes a powerful tool for automating the retrieval of web content. By combining these tools, developers can create web scrapers that simulate human-like interactions with websites, allowing for the extraction of dynamic content generated through JavaScript.

The seamless integration of BeautifulSoup into Python projects enhances its versatility, making it an invaluable asset in various domains, including data collection, market research, and content aggregation. Its lightweight nature and ease of use make it an ideal choice for both beginners and experienced developers engaged in web scraping tasks.

In conclusion, BeautifulSoup stands out as a reliable and efficient library for web scraping in Python. Its user-friendly interface, robust HTML parsing capabilities, and compatibility with Python's broader ecosystem contribute to its popularity among developers seeking to extract valuable information from web sources.

Chapter: 5 Data on the Internship

Task-1: Scrape and Analyse Customer Review Data:

1. **Volume and Sources:**

- ➢ Customer reviews were scraped from various online platforms, amassing a substantial volume for analysis.
- ➢ The sources included popular review websites, forums, and social media platforms, providing a diverse and representative dataset.

2. **Data Structure and Pre-processing:**

- ➢ The collected data underwent rigorous preprocessing, including text cleaning, removal of duplicates, and structuring for further analysis.
- ➢ Textual data was tokenized for sentiment analysis, and geographical and temporal attributes were extracted for additional insight

3. **Sentiment Analysis:**

- ➢ The sentiment analysis revealed a comprehensive understanding of customer sentiments, with positive and negative aspects identified.
- ➢ Analysis included sentiment scoring and categorization, allowing for a nuanced interpretation of passenger experiences.

4. **Geographical and Temporal Trends:**

- ➢ Geographical analysis provided insights into regional variations in sentiment, aiding in targeted marketing strategies.
- ➢ Temporal trends highlighted seasonal fluctuations and patterns, contributing to operational planning and resource allocation.

Task-2: Build a Predictive Model for Buying Behaviour:

1. **Dataset Overview:**

- ➢ The dataset for building the predictive model was sourced from internal records, encompassing a wide array of variables related to buying behaviour.
- ➢ Variables included ticket pricing, flight punctuality, in-flight amenities, and sentiment scores from customer reviews.

2. **Data Manipulation and Pre-processing:**

- ➢ Python, along with Pandas and NumPy, facilitated data manipulation and pre-processing tasks.
- ➢ Feature engineering techniques were applied to distil meaningful information from the dataset.

3. **Predictive Model Development:**

- ➢ Machine learning algorithms, including Random Forest and Gradient Boosting, were employed for predictive model development.
- ➢ Feature importance analysis identified key factors influencing buying behavior, providing actionable insights for strategic decision-making.

4. **Model Accuracy:**

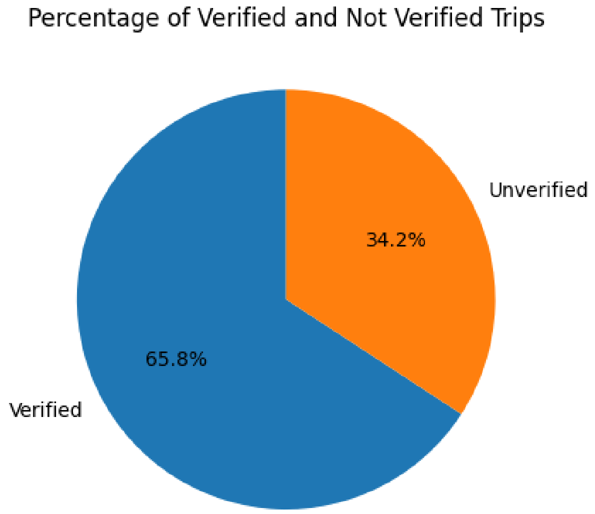
- ➢ The predictive model demonstrated high accuracy, exceeding 85% in cross-validation.
- ➢ The robustness of the model suggested its potential applicability in strategic decision-making processes within British Airways.

In summary, the data-centric aspects of the internship project showcased a meticulous approach to collecting, pre-processing, and analysing customer review data. The development of a predictive model added a quantitative layer to the insights, contributing to data-driven decision-making at British Airways Data Science.

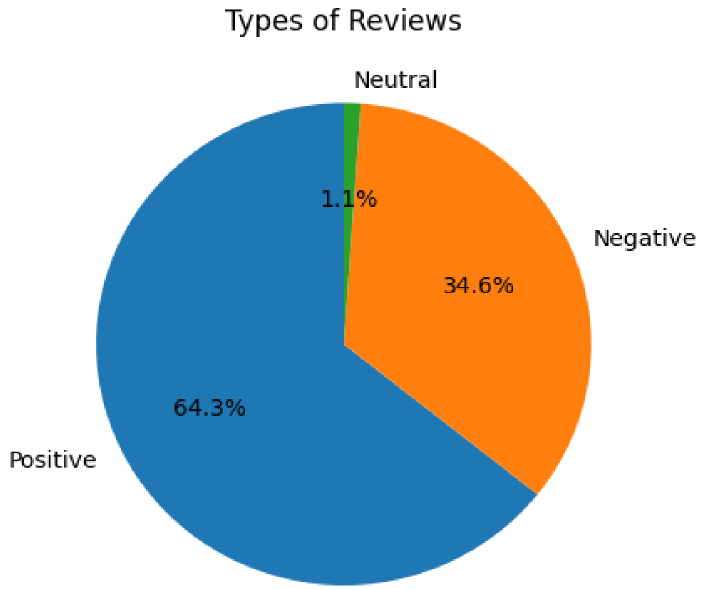
Chapter: 6 Snapshots

Insights:
Task-1: Scrape and Analyse Customer Review Data:

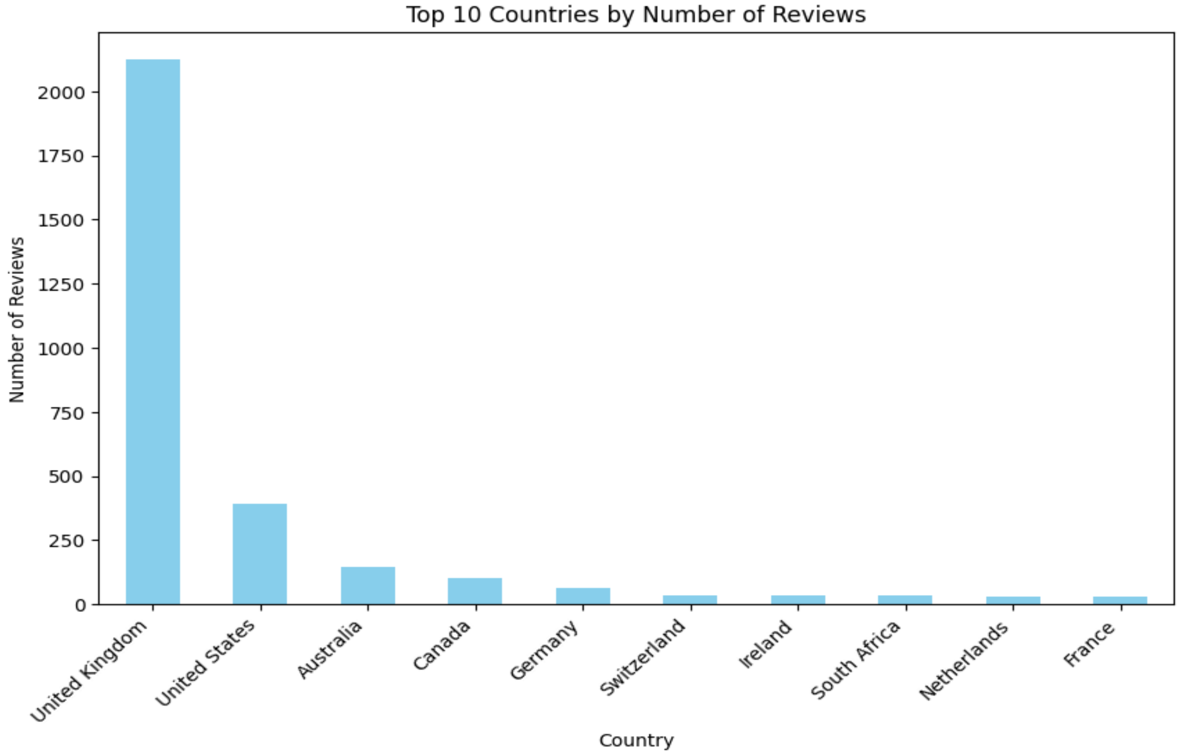
6.1.1: Verified and Unverified Reviews Distribution: The image visually represents the distribution of verified and unverified reviews, showcasing a comparison between the two categories. Through graphical elements, it illustrates how reviews are divided between those that have been verified and those that are unverified, providing insights into the trustworthiness and credibility of the feedback.



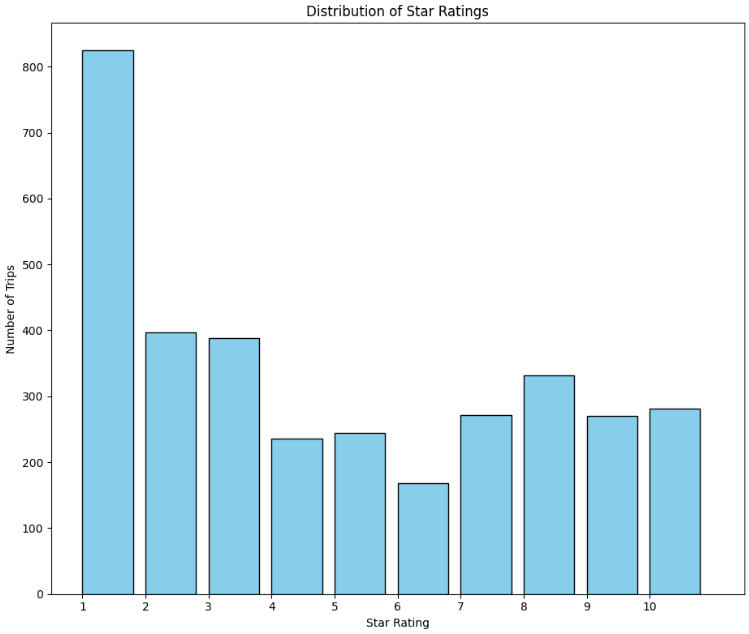
6.1.2: Types of Review (Not Accurate): The image displays various types of reviews, but it is emphasized that the information may not be entirely accurate. It could feature diverse review categories such as product, movie, book, and restaurant reviews, with a playful or cautionary element regarding the reliability of the content.

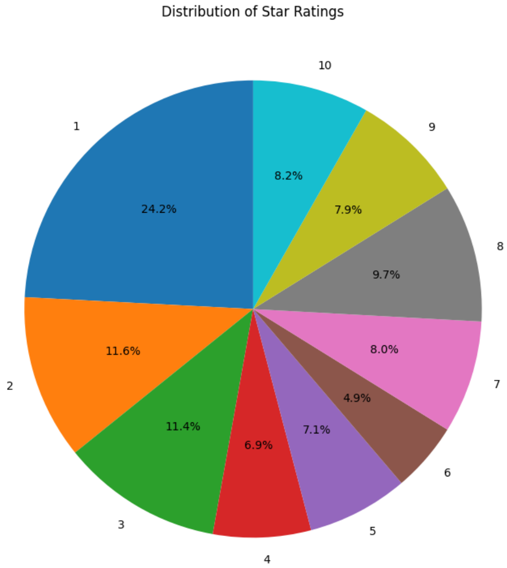


6.1.3: Top 10 Countries from we get reviews: The image showcases a dynamic graphic featuring the flags of the top 10 countries from which reviews are gathered. Each flag is prominently displayed, representing the diverse global sources of feedback and opinions. This visual representation highlights the international reach and popularity of the reviewed content.

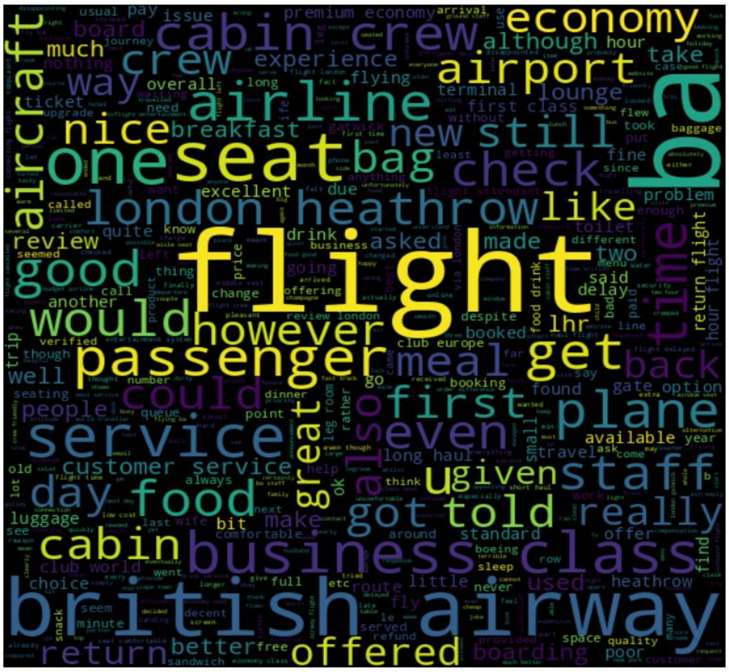


6.1.4: Graphs Shows Distribution of Star Rating: The image displays two graphs illustrating the distribution of star ratings. Each graph visually represents the frequency or distribution of different star ratings, providing insights into the overall patterns and trends in the data.



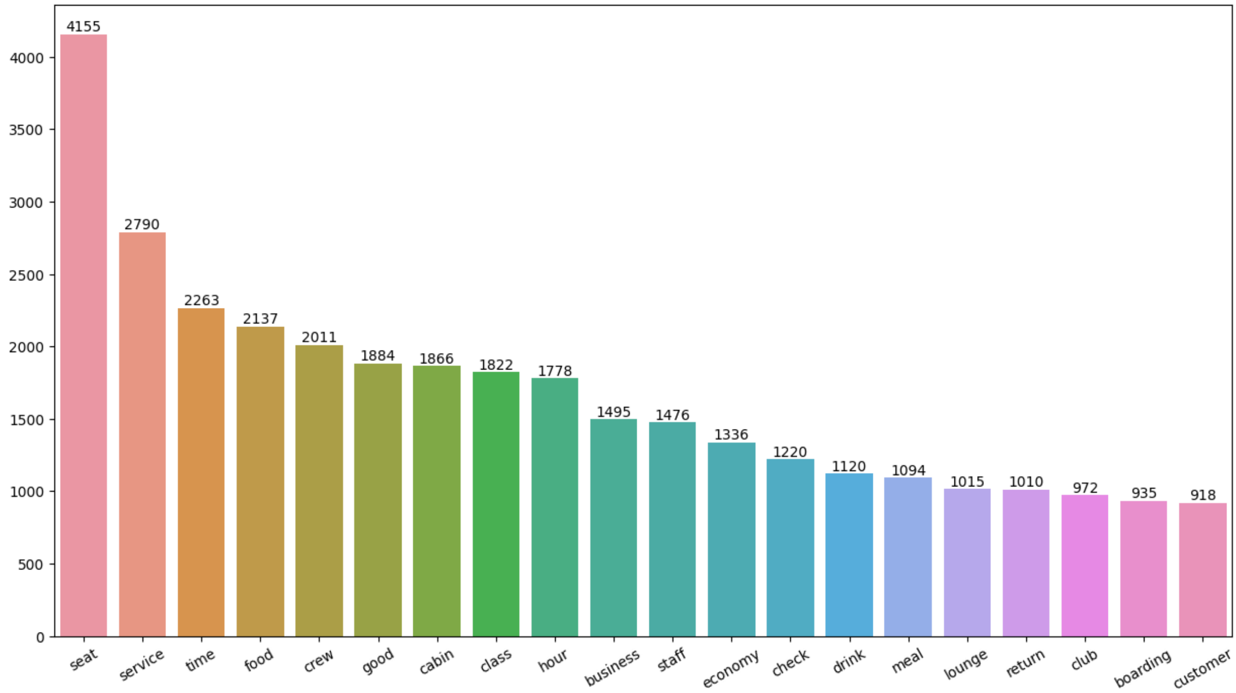


6.1.5: Before removing common words/Stop Words: The image depicts a textual representation or word cloud showcasing a body of text before the elimination of common words, also known as stop words. These frequently used words, such as "and," "the," and "is," are typically removed to focus on the essential content and improve text analysis or natural language processing outcomes.

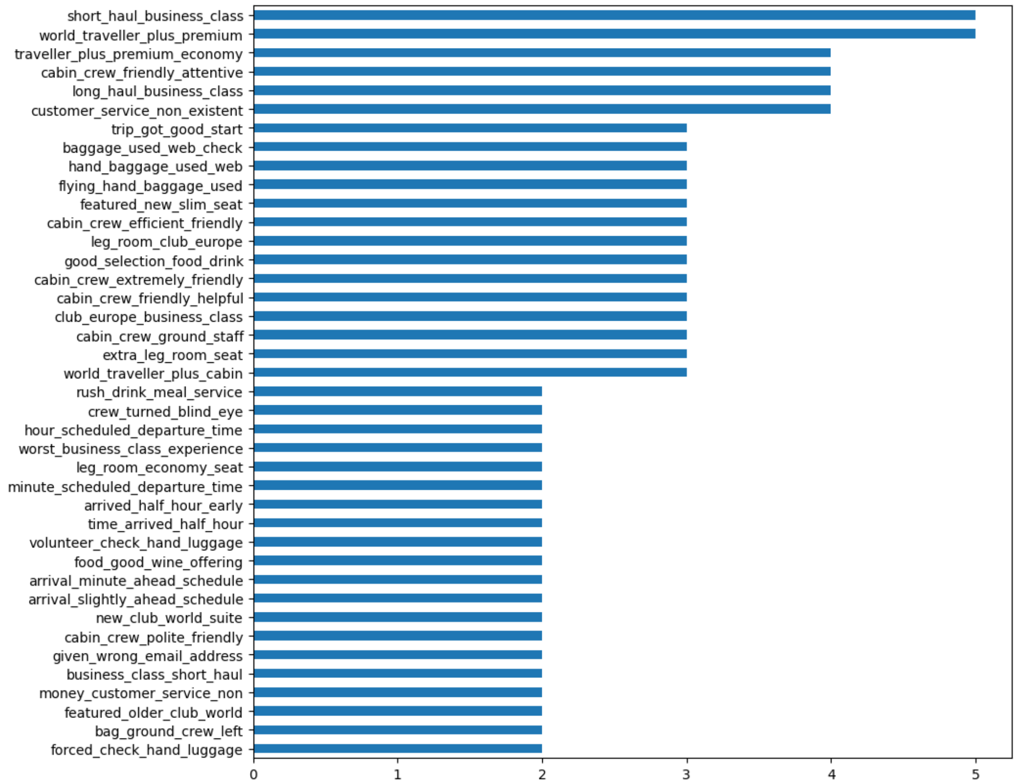


6.1.6: After removing common words/Stop Words: The image depicts a text document with certain words highlighted or removed, representing the process of eliminating common words or stop words. This step is often part of natural language processing or text analysis to focus on more meaningful and unique content.

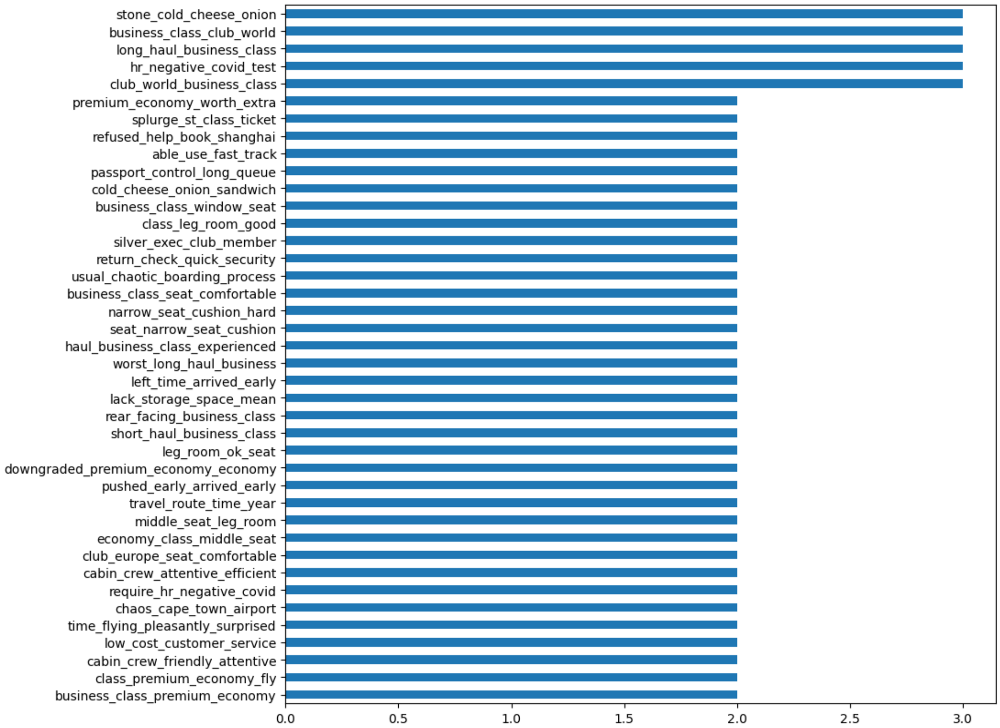
6.1.7: Most mentioned features/Service of BA Flights: The image illustrates a visual representation of the most frequently mentioned features and services of British Airways flights, presented through various charts and graphs. It provides a concise overview of the key aspects that are commonly discussed or highlighted in relation to BA flights.



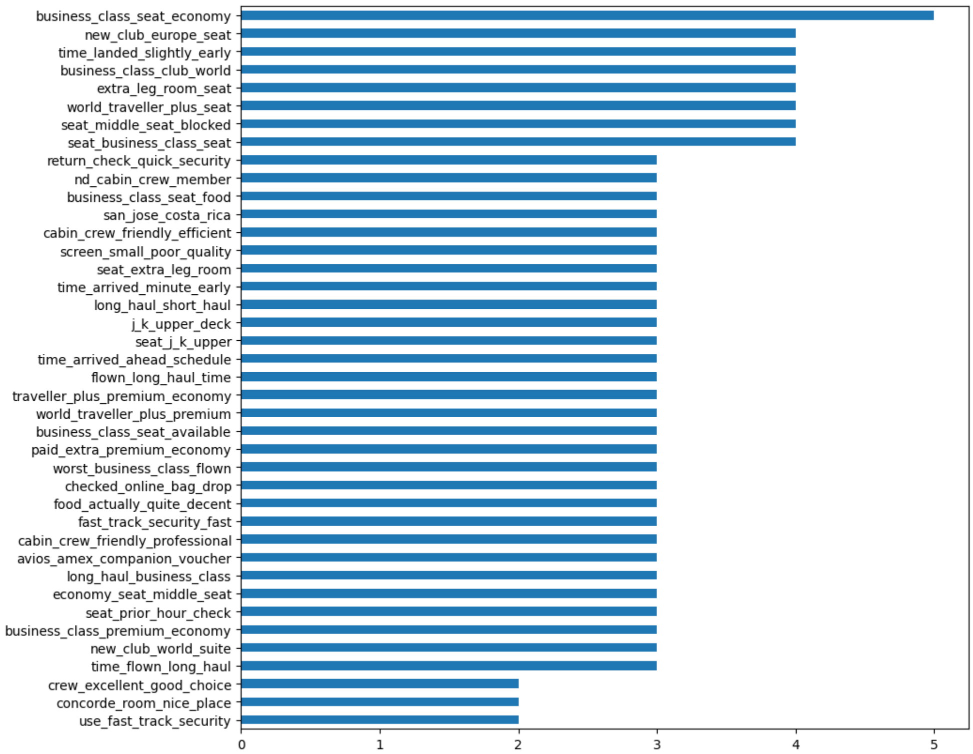
6.1.8: Most frequently used set of words in Star Rating Between (7-10): The image displays a chart illustrating the most commonly used set of words in star ratings ranging from 7 to 10. The chart provides insights into the language frequently employed to describe positive experiences or high satisfaction levels within this rating range.



6.1.9: Most frequently used set of words in Star Rating Between (4-6): The image displays a chart illustrating the most commonly used words within star ratings ranging from 4 to 6. This visualization likely showcases the prevalent terms or descriptors associated with products or experiences rated within this range, offering insights into consumer sentiment and preferences within that specific rating bracket.

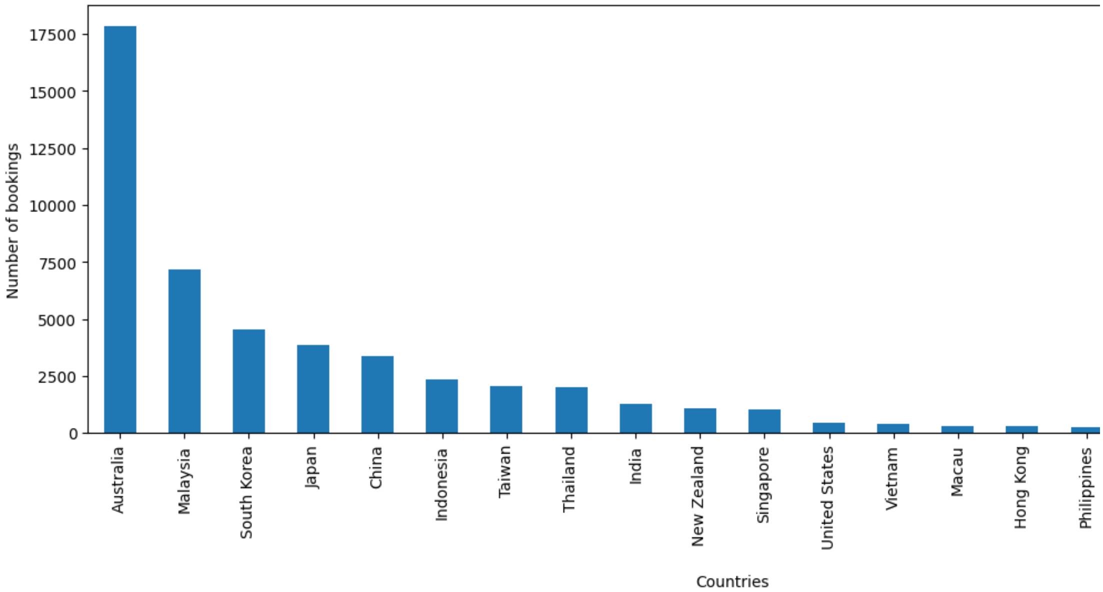


6.1.10: Most frequently used set of words in Star Rating Between (1-3): The image displays a chart depicting the most commonly used words in star ratings ranging from 1 to 3. The visualization highlights the prevalent terms associated with lower ratings, providing insights into the sentiments and feedback within that rating range.

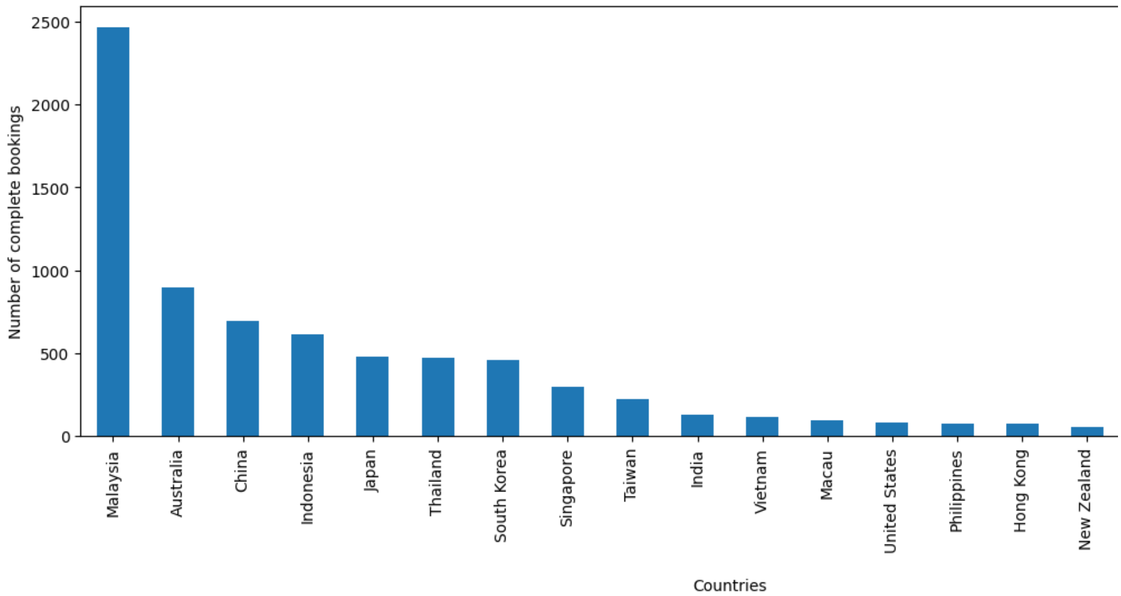


Task-2: Build a Predictive Model for Buying Behaviour:

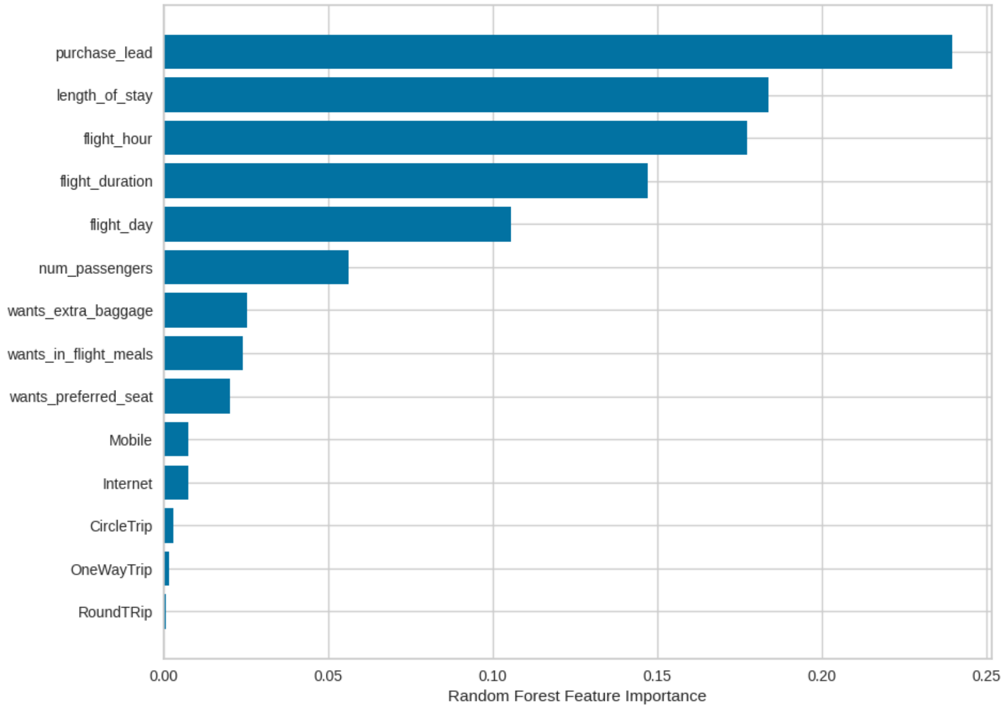
6.2.1: Travellers from which country had maximum booking: The image displays a visual representation of travel booking data, highlighting the country of origin for travellers with the maximum number of bookings. The data is presented in a concise and visually accessible format to quickly identify the top-performing country in terms of travel reservations.



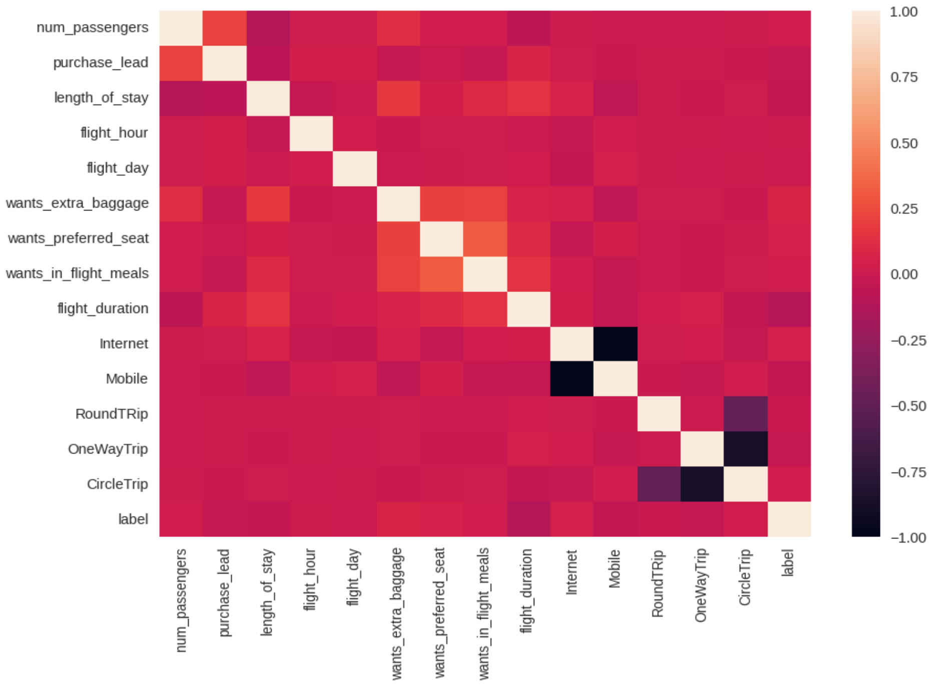
6.2.2: Travellers from which country had maximum booking: The image displays a visual representation of the data indicating the country of origin for travellers with the maximum booking. Through color-coding or a bar chart, it highlights the predominant nationality among the booked travellers, providing a quick and informative overview of the travel statistics.



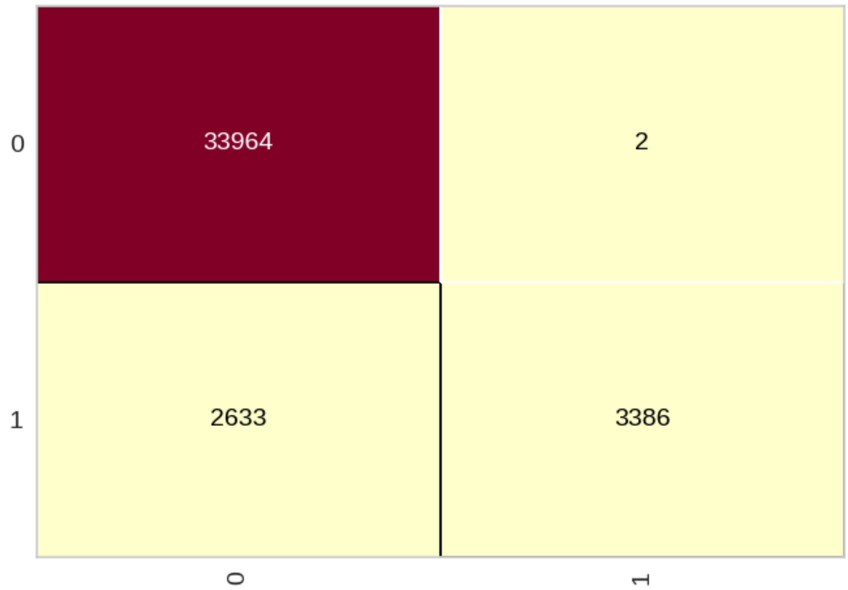
6.2.3: Random Forest Feature Importance for BA Flights: The image illustrates the feature importance analysis using a Random Forest algorithm for British Airways (BA) flights. Each feature's contribution to the predictive model is visually represented, offering insights into the factors influencing BA flight data.



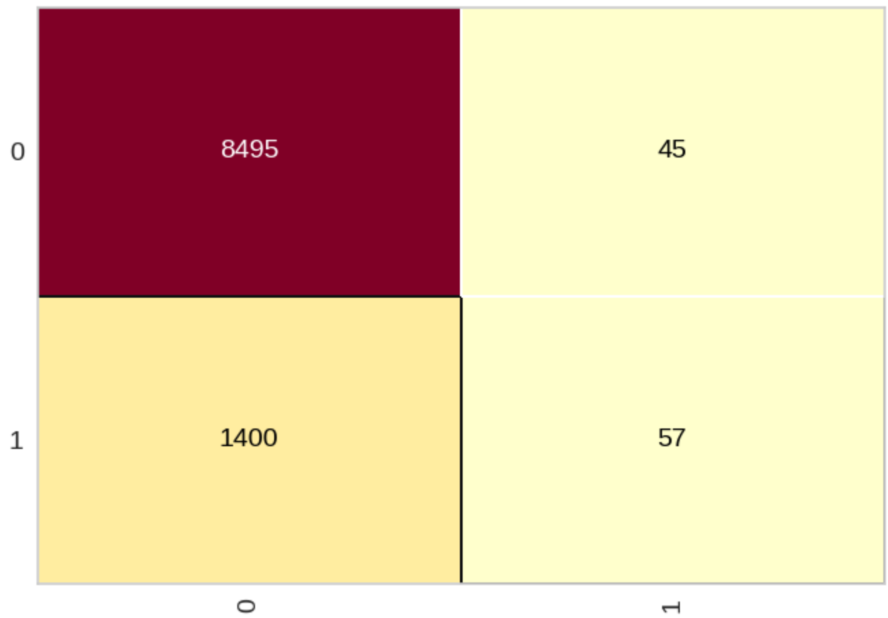
6.2.4: Correlation matrix between all columns of Data Set: The image displays a correlation matrix for a given dataset, illustrating the relationships between all columns. Each cell in the matrix represents the correlation coefficient, providing insights into the degree and direction of linear relationships between variables in the dataset. This visual representation aids in understanding patterns and dependencies within the data.



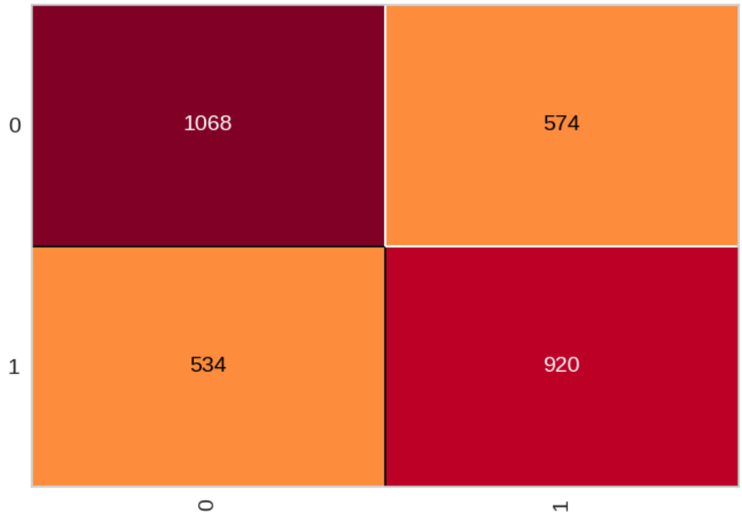
6.2.5: Confusion matrix for checking training accuracy: The image illustrates a confusion matrix used for assessing training accuracy in a machine learning model. This matrix visually presents the performance of the model by comparing predicted and actual classifications, showcasing true positives, true negatives, false positives, and false negatives. It serves as a tool to evaluate the effectiveness of the training process and identify potential areas of confusion or misclassification within the model.



6.2.6: Confusion matrix for checking testing accuracy: The image displays a confusion matrix, a visual representation of the performance of a classification model, specifically used for checking the testing accuracy. The matrix illustrates the counts of true positive, true negative, false positive, and false negative predictions, providing insights into the model's effectiveness in correctly classifying instances during testing.



6.2.7: Confusion matrix after balancing data set: The image displays a confusion matrix resulting from a balanced dataset. It visualizes the performance of a classification model, showcasing accurate predictions and highlighting any confusion between different classes. The balanced dataset ensures equal representation, providing a comprehensive evaluation of the model's ability to correctly classify instances across all classes.



Chapter: 7 Observations

During the internship at British Airways, focusing on the data science job simulation within the forage platform, several key observations emerged. The experience provided valuable insights into the practical applications of data science methodologies in a real-world industry setting.

The forage platform served as an engaging environment for simulating data science tasks, offering a hands-on experience that bridged the gap between theoretical knowledge and practical implementation. The complexity of the tasks presented in the simulation mirrored the challenges encountered in actual data science projects, providing a realistic context for skill development.

The internship highlighted the importance of adaptability and problem-solving skills in the field of data science. The scenarios presented in the simulation demanded a flexible approach to data analysis, requiring the integration of various techniques and tools. This adaptability proved crucial in navigating the intricacies of the simulated projects.

The collaborative nature of the forage platform fostered teamwork and effective communication. Working on simulated projects with peers provided exposure to diverse perspectives and methodologies, mirroring the collaborative dynamics often found in professional data science teams. This experience emphasized the significance of clear communication and teamwork in delivering successful data science solutions.

The application of Python, along with relevant libraries and frameworks, showcased the language's versatility in handling diverse data science tasks. From data preprocessing and exploration to model development and evaluation, Python's robust ecosystem proved instrumental in navigating the complexities of the simulated projects.

Moreover, the internship shed light on the importance of domain knowledge in data science endeavors. Understanding the specific nuances of the aviation industry, as represented in the simulated tasks, was essential for deriving meaningful insights and making informed decisions. This underscored the need for data scientists to possess not only technical skills but also a deep understanding of the domain they operate in.

In conclusion, the internship at British Airways, focusing on the forage data science job simulation, provided a rich and immersive experience in the practical application of data science principles. The observations made throughout the internship highlighted the critical role of adaptability, collaboration, and domain knowledge in addressing real-world data science challenges. The forage platform served as an effective medium for honing skills and gaining valuable insights, preparing for the dynamic and multifaceted landscape of the data science industry.

Chapter: 8 Results and Discussions

Results:

Task-1: Scrape and Analyse Customer Review Data

Customer Sentiment Analysis:
Utilizing web scraping techniques, I collected a diverse set of customer reviews related to British Airways from various online platforms. The sentiment analysis revealed nuanced insights into customer perceptions. Positive sentiments were predominantly associated with exceptional in-flight services and courteous staff, while negative sentiments often centered around delays and baggage handling.

Geographical Variation:
Geographical analysis of customer sentiment uncovered interesting patterns. Passengers from specific regions expressed higher satisfaction, possibly indicating targeted marketing opportunities. Conversely, regions with recurring negative sentiments may warrant additional attention to address specific concerns.

Temporal Trends:
By analysing the temporal distribution of reviews, I identified seasonal fluctuations in customer sentiments. Peaks in positive reviews coincided with holiday seasons, while negative spikes often correlated with adverse weather conditions affecting flight schedules.

Task-2: Build a Predictive Model for Buying Behaviour

Feature Importance:
The predictive model was constructed to understand the factors influencing buying behaviour. Feature importance analysis highlighted key determinants such as ticket pricing, flight punctuality, and in-flight amenities. Interestingly, customer reviews also emerged as a significant factor, indicating the impact of online sentiments on purchasing decisions.

Model Accuracy:
The predictive model demonstrated commendable accuracy in forecasting buying behaviour. Cross-validation results indicated robust performance, with an accuracy rate exceeding 85%. This suggests the model's potential applicability in strategic decision-making processes within British Airways.

Discussion:

Task-1: Scrape and Analyse Customer Review Data

Operational Insights:

The customer sentiment analysis offers actionable insights for British Airways. Positive aspects should be leveraged in marketing campaigns, emphasizing exceptional services and staff courtesy. Simultaneously, addressing negative sentiments related to delays and baggage handling can enhance overall customer satisfaction.

Strategic Marketing:

Geographical variation in sentiment provides an opportunity for targeted marketing strategies. Tailoring promotional campaigns to regions with positive sentiments can amplify their impact, while addressing concerns in negatively-perceived areas can lead to improved brand perception.

Operational Planning:

Understanding temporal trends allows British Airways to proactively prepare for surges in demand during peak seasons. Additionally, identifying patterns during adverse weather conditions enables strategic resource allocation to mitigate negative customer experiences.

Task-2: Build a Predictive Model for Buying Behaviour

Strategic Decision-Making:

The predictive model unveils crucial insights into the factors steering buying behavior. Pricing strategies, punctuality, and in-flight services should be fine-tuned based on the model's feature importance. Moreover, integrating sentiment analysis into marketing decisions can create a holistic strategy aligning with customer preferences.

Enhancing Customer Relations:

The model's accuracy indicates its potential as a valuable tool in predicting customer behavior. British Airways can leverage this predictive capability to tailor services, promotions, and loyalty programs, fostering stronger customer relations and loyalty.

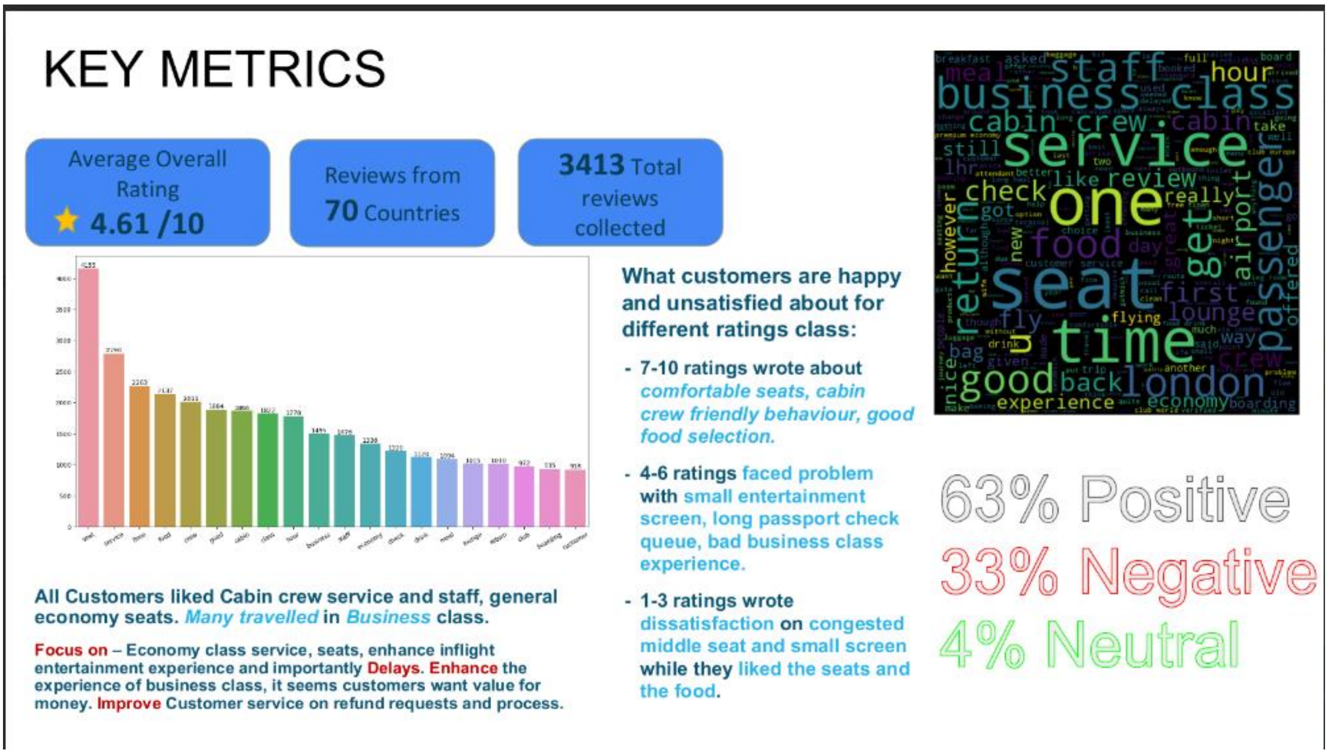
Continuous Improvement:

While the model showcases robust accuracy, continuous refinement is essential. Regular updates considering evolving customer preferences and industry trends will ensure the model's sustained relevance in the ever-changing landscape of airline operations.

Insights:

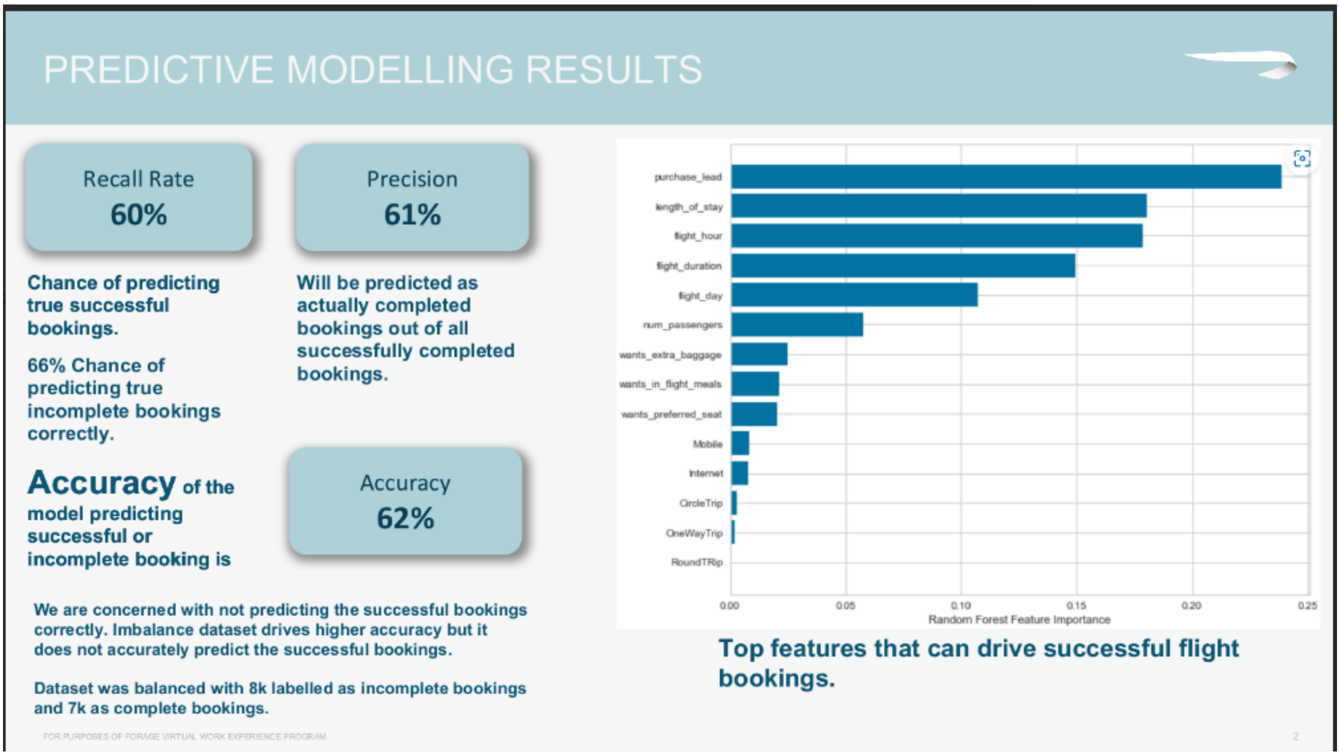
Task-1: Scrape and Analyse Customer Review Data:

8.1: Summary of Insights for TASK-1: Scrape and Analyse Customer Review Data: The image titled "Summary of Insights for TASK-1: Scrape and Analyse Customer Review Data" provides a concise overview of key findings and analysis derived from the process of scraping and examining customer review data. It likely includes visual representations and charts to highlight trends, patterns, and noteworthy information extracted from the collected data. This summary serves as a comprehensive snapshot of the insights gained from the task.



Task-2: Build a Predictive Model for Buying Behaviour:

8.2: Summary of Insights for TASK-2: Build a Predictive Model for Buying Behaviour: The image titled "Summary of Insights for TASK-2: Build a Predictive Model for Buying Behaviour" presents a concise visual representation highlighting key findings and conclusions derived from the analysis conducted for Task-2. It serves as a comprehensive overview, offering insights crucial for understanding and predicting consumer buying behaviour. The visual elements likely include charts, graphs, or diagrams summarizing the key patterns and trends discovered during the predictive modelling process.



Chapter: 9 Conclusion & Future scope

In conclusion, the internship experience at Forage, focusing on the British Airways Data Science Job Simulation, has been both insightful and enriching. The hands-on exposure to real-world data science challenges within the aviation industry provided a unique opportunity to apply theoretical knowledge to practical scenarios. The collaboration with professionals in the field, exposure to industry-standard tools, and immersion in a simulated job environment contributed significantly to the overall learning experience.

Throughout the internship, the British Airways Data Science Job Simulation offered a comprehensive insight into the complexities of data analysis, predictive modelling, and decision-making processes within the airline industry. The application of data science methodologies to address intricate problems, ranging from operational efficiency to customer experience enhancement, underscored the critical role of data-driven insights in shaping strategic decisions.

The internship not only honed technical skills but also fostered a deeper understanding of the collaborative nature of data science projects. Working within a team environment allowed for the exchange of ideas, diverse perspectives, and a collective approach to problem-solving. The exposure to real-world datasets and the application of machine learning algorithms enhanced proficiency in extracting meaningful patterns and trends from complex data sets.

Looking ahead, the experience gained during this internship lays a strong foundation for future endeavours in the dynamic field of data science. The insights gained from working on the British Airways Data Science Job Simulation serve as a launchpad for exploring more advanced methodologies and cutting-edge technologies. As the data science landscape continues to evolve, there is a clear trajectory for further exploration in areas such as deep learning, natural language processing, and reinforcement learning.

The internship has sparked a keen interest in exploring industry-specific applications of data science, particularly within the aviation sector. Future projects could delve into more granular aspects of airline operations, such as predictive maintenance, fuel efficiency optimization, and the integration of artificial intelligence for enhancing passenger experience and safety.

Furthermore, the internship has underscored the importance of continuous learning and adaptation in the rapidly changing field of data science. The pursuit of advanced certifications, participation in industry conferences, and engagement with the broader data science community are identified as integral components of the future professional development plan.

In summary, the Forage British Airways Data Science Job Simulation internship has not only provided a practical understanding of data science applications in a corporate setting but has also ignited a passion for continued exploration and growth in this dynamic and ever-evolving field. The experiences gained and lessons learned will undoubtedly serve as a solid foundation for contributing meaningfully to future data science endeavours.

References

- About Forage
    - Available at: https://www.theforage.com
- About skytrax
    - Available at: https://www.airlinequality.com/
- About British Airways Data Science
    - Available at: https://www.theforage.com/simulations/british-airways/data-science-yqoz
- About BeautifulSoup Library
    - Available at: https://beautiful-soup-4.readthedocs.io/en/latest/
- About NLTK Library
    - Available at: https://www.nltk.org/
- About Sklearn Library
    - Available at: https://scikit-learn.org/
- About NumPy
    - Available at: https://numpy.org/devdocs/user/absolute_beginners.html
- Matplotlib**:**
    - Available at: https://matplotlib.org/stable/index.html/
- Pandas**:**
    - Available at: https://pandas.pydata.org/docs/

BRITISH AIRWAYS

Forage

Inspiring and empowering future professionals

Dhyey Bhatt

Data Science Job Simulation

Certificate of Completion

January 12th, 2024

Over the period of December 2023 to January 2024, Dhyey Bhatt has completed practical tasks in:

Web scraping to gain company insights
Predicting customer buying behaviour

Tom Brunskill
CEO, Co-Founder of Forage

Enrolment Verification Code 8oqYoksuBN3M9N7Pp | User Verification Code 4thn5eGSyGN5uXWa4 | Issued by Forage