A black background with grey leaves

Description automatically generated

*A grey logo on a black background

Description automatically generated*

***Designing of a Data Analytics Framework For Strategic Decision Making in the Hospitality Sector***

A close-up of a logo

Description automatically generated

***SUBMITTED***

***BY***

***ANKITHA***

***( )***

*A dissertation submitted in partial fulfilment of the requirements for* ***MSc Advanced Computer Science*** *by taught program, supervised by Professor* ***Edimore Chikihore.***

**CONTENTS**

**ABSTRACT ………………………………...……….………………………………….……………………….. 3**

**ACKNOWLEDGMENTS .…………………………...………………………………………...…..................... 5**

**INTRODUCTION/BACKGROUND ..………………………...……………………………………………… 6**

**LITERATURE REVIEW .………………………………………………...………………………...,………… 8**

**AIMS, OBJECTIVES AND RESEARCH QUESTIONS …………………………………………………..... 12**

**METHODOLOGY AND IMPLEMENTATION** …….…………………………...………………………...... **13**

1. **PROBLEM STATEMENT ……………………………………………………………………………... 14**
2. **DATASET DESCRIPTION …………………………………………………………………………….. 15**
3. **EDA USING PYTHON …………………………………………………………………………………. 16**
4. **UNDERSTANDING DATA USING MSSQL DATABASE ………..………………………………….. 17**
5. **DASHBOARD CREATION USING POWERBI ……………………………………………………... 18**

**RESULTS .……………………………………………………………………………………...………………. 19**

**DISCUSSION AND CONCLUSION .………………………………………………………………………… 20**

**REFERENCES** **………………………………………………………………………………….................… 21**

**APPENDICES** **…………………………………………………………………................................. 22**

**ABSTRACT**

In an era where data reigns supreme, the hospitality sector stands at a pivotal juncture, poised to harness its full potential to drive strategic decision-making. For hospitality businesses to prosper in a highly competitive environment, building and implementing a strong data analytics platform is both an opportunity and a requirement. Specifically designed for the hospitality industry, our suggested data analytics framework incorporates state-of-the-art tools and techniques to provide decision-makers with making useful decisions. A thorough grasp of consumer behaviour, industry trends, and operational effectiveness is made possible by this framework, which covers data collection, data exploration and, analysis carried out by emerging tools like Python and SQL followed by visualization using business intelligence tools.

**Key components of our framework include:**

**Data Integration and Collection:** - Aggregating data from diverse sources such as reservations systems, guest feedback platforms, social media, and IoT devices to create a unified data repository.

**Real-time Monitoring: -** Implementing real-time monitoring dashboards using a business intelligence tool to track key performance indicators (KPIs), identify anomalies, and facilitate best insights for making decisions.

**Strategic Insights: -** Generating actionable insights through interactive data visualizations and intuitive dashboard reports, enabling stakeholders/consumers to uncover opportunities, mitigate risks, and drive innovation.

**Scalability and Flexibility: -** Designing a scalable and flexible architecture that accommodates evolving business requirements, technological advancements, and regulatory compliance.

By adopting our data analytics framework, hospitality organizations can realize several strategic benefits:

**Enhanced Guest Experiences**: Personalized offerings and tailored services based on deep insights into guest preferences and behaviour.

**Operational Efficiency:** Streamlined processes, optimized resource allocation, and cost-effective operations.

**Competitive Advantage:** Data-driven decision-making enables initiative-taking responses to market trends, ensuring a competitive edge in a dynamic industry landscape.

**Revenue Growth:** Improved pricing strategies, upsell opportunities, and targeted marketing initiatives drive revenue growth and profitability.

**Risk Mitigation:** Early identification of potential risks and initiative-taking mitigation strategies safeguard the reputation and integrity of the brand.

The implementation of a data analytics framework tailored to the unique needs of the hospitality sector is imperative for organizations seeking to thrive in an increasingly competitive and digitally driven environment. By harnessing the power of data, hospitality enterprises can unlock new avenues for growth, innovation, and excellence in guest satisfaction.

**ACKNOWLEDGEMENTS**

I would like to sincerely thank my supervisor Edmore Chikohora for his timely advice, help, and direction during the dissertation.

**INTRODUCTION**

In the dynamic and fiercely competitive landscape of the hospitality sector, understanding customer preferences, optimizing operations, and making informed decisions are paramount to success. The industry's vulnerability to shifting consumer tastes, market trends, and global events necessitates a departure from conventional decision-making methods towards more sophisticated approaches. Leveraging data analytics has emerged as a crucial strategy for hospitality organizations seeking to stay ahead in this rapidly evolving environment. With the exponential growth of data within the hospitality industry, there lies an opportunity to extract invaluable insights through meticulous analysis. This research proposal aims to explore the transformative potential of integrating Excel, SQL, Python, and using business intelligence tools like PowerBI for strategic decision-making within the hospitality domain. By harnessing the strengths of each tool, this study seeks to provide a comprehensive framework that empowers industry practitioners to navigate complexity, drive sustainable growth, and enhance guest satisfaction.

**BACKGROUND**

The hospitality industry encompasses a wide array of sectors including lodging, dining, travel, and tourism, each subject to the whims of consumer behaviour and global dynamics. Traditional decision-making methodologies are proving inadequate in the face of these challenges, necessitating a paradigm shift towards data-driven strategies. At the core of this shift lies the integration of data analytics tools such as SQL, Python, and Power BI, offering a potent arsenal for extracting actionable insights from vast datasets.

**PYTHON and Structured Query Language (SQL)** serves as the backbone for managing and manipulating large datasets within hospitality information systems. Its ability to facilitate efficient data retrieval and pattern recognition empowers decision-makers to glean valuable insights from historical data, guiding strategic choices in an ever-changing landscape. **Microsoft Excel,** a ubiquitous tool in business analytics, provides an intuitive platform for data exploration, descriptive analytics, and dashboard development. Its accessibility makes it an ideal candidate for integration into existing decision-making processes within the hospitality industry. In this project we are not exploring the insights using the excel. The dataset are available in the form of Excel (csv files) these excel files are loaded to Google Collab for exploring data analysis and visualisation, Exporting these datasets to a database and on other hand these excel csv files are loaded to POWERBI and will perform data analysis and exploration by using POWEBI functionalities like Power query and Dax functions and creates key insights by creating a dashboard. **POWERBI** Meanwhile, business intelligence software like Power BI offers dynamic dashboards that amalgamate data from multiple sources, providing decision-makers with real-time analytics and a holistic view of key performance indicators. This enables data-driven collaboration and supports informed decision-making at every level of the organization.

**LITERATURE REVIEW**

One of the biggest and fastest-growing industries in the world is international hospitality and tourism. Due to its anticipated steady growth, it currently generates US$7.6 trillion in revenue and 292 million jobs globally, or one in ten jobs and nearly 10% of the global GDP (World Travel & Tourism Council, 2017; Statista Research Department, 2018; OECD, 2018).

In an effort to stand out from the competition and win over new clients, service companies are increasingly managing customer experiences. According to **Zomerdijk and Voss (2010)**, the service offering and design revolve around the concept of service experience. Customer experiences are ever-changing.

According to a PWC analysis, the hotel sector will increase by 3.9% on average over the next ten years. Additionally, 86% of consumers are willing to spend more for a better customer experience, but only 1% of suppliers of goods and services will be able to live up to these expectations **(Puthiyamadam & Reyes, 2020).**

According to **Gilmore and Pine (2002)**, the tourism and hospitality sectors have historically been categorized as some of the more experience-based service sectors. Customers' behaviour is influenced by their perception of a website, which also significantly affects their intention to return or make additional purchases.

A small number of people have also developed web quality using the Technology Acceptance Model and Theory of Reasoned Action **(Loiacono et al. 2007).** If clients have the intended favourable experience, they also become loyal customers, which boosts revenues. Improving the customer experience is one of the areas on which businesses are focusing more and more. Numerous scholars have debated the conceptual distinction between service quality and customer experience **(Klaus & Maklan, 2013)** and created new metrics to quantify it (**Klaus & Maklan, 2012).**

In his "Digital Service Usage Satisfaction Model," proposed by **Kar (2020),** the author methodologically aims to validate a novel technique that leverages social media data to construct an inferential theoretical model. The goal of the study was to forecast the usage satisfaction of mobile payments by extending and tying together different theories of technology adoption with literature on service quality.

Big data and social media's widespread use today suggest that businesses have access to far more consumer information than they did in the past **(Wilson, 2018). B**

According to a Robert Half report on the "Future of Work," 39% of IT leaders currently use artificial intelligence (AI) or machine learning, 33% anticipate using it in the next three years, and 19% anticipate using it in the next five years **(Robert Half, 2020).** In contrast to conventional data analytics software, artificial intelligence (AI) can predict customer behaviour and is always learning from the data it analyses. This enables brands to enhance the customer journey, boost sales opportunities, and deliver extremely relevant content **(Clark, 2021).** From the levels of personalization that are currently common in all service industries, this is a step towards hyper personalization.

With contemporary AI systems, transparency can be particularly challenging, particularly with those built on deep learning systems. In addition, privacy and transparency clash because individuals using machine learning models for training may not want their data or conclusions drawn from it to be public. One feature of a typical machine learning algorithm, such as an artificial neural network (ANN), is that it is nearly impossible to look inside the ANN after it has been trained with datasets in order to figure out how and why it makes a given decision. These systems are known as "black-boxes," and numerous scholars have maintained that they need to provide justifications for their actions **(Selbst & Barocas, 2018).**

The development of technology has led to an increase in the use of social media for idea and information sharing. TF-IDF and Bag of Word can be used to create sentiment-based analysis models. "Term Frequency — Inverse Document Frequency" is what TF-IDF stands for. This method counts the number of words in a collection of documents. Additionally, different machine learning approaches could,

The latest technological developments have an effect on all businesses. These developments include new methods for gathering large amounts of data, like cloud computing, and sophisticated analytics for deriving insights from the data, like machine learning (ML) **(Khan, Garg & Rahman, 2015)**. But when business operations get more intricate and the volume of data processed rises, major problems appear. Businesses still lack insight even in the face of an abundance of data **(Bolton et al., 2018).**

With the help of machine learning (ML), systems can automatically learn from experience without the need for human programming. Following data analysis, the learning is immediately put into practice to improve the procedure **(Hossain, Akter, Kattiyapornpong & Wamba, 2017).**

Machine learning is one of the preferred analytics modes for big data processing. Due to recent advancements in computing skills that have significantly decreased the cost of its algorithms, this area of AI has become more well-known (**Nilashi, Bagherifard, Rahmani & Rafe, 2017).** According to recent studies, machine learning will soon completely change the nature of the market. Its wide range of options has been shown to disrupt labor and force businesses to redefine their marketing strategies (**Juaneda-Ayensa, Mosquera & Sierra Murillo, 2016**).

Conventional statistical models impose a number of strict limitations on the kinds of data and how they are distributed, which may reduce the analysis's potential **(Cui, Wong & Lui, 2006).** The main presumptions of the research methods are frequently broken, which causes some problems for the models when they are applied to actual data. As a result, the laborious and time-consuming procedures required to create a knowledge-based system were eliminated by the development of machine learning (ML) technologies **(Bose and Mahapatra, 2001).**

In classical economics, the consumer is seen as a critical decision-maker who bases their choices on reason. However, contemporary marketing theories contend that irrational and intangible factors have a strong correlation with consumers' perceived worth (**Babin, Darden & Griffin, 1994).** Because it encompasses all aspects of a business's service, customer experience is crucial **(Reis, Ruivo, Oliveira & Faroleiro, 2020).**

Customer experience is a multifaceted construct with cognitive, emotional, behavioural, sensory, and social components, according to academics and business owners (**Zollo, Filieri, Rialti & Yoon, 2020**).

Because there are so many points of contact between a company and its clients, it has recently become clear from the analysis of the interactions those points generate that the business is paying attention to its clients. Consequently, it is more crucial than ever to consider aspects of emotional and irrational consumer behaviour (**Vošner, Kokol, Bobek, Železnik & Završnik, 2016**).

A triumph contingent on the calibre of services and contentment of patrons provided by establishments engaged in hospitality and tourism. Aside from cost, dependability, and strategically placed highly developed destinations, drawing and keeping visitors is primarily dependent on the calibre of the services provided and the level of customer satisfaction (**Narayan et al., 2008; Bellou and Andronikidis, 2009; Nunkoo et al., 2017; Li et al., 2020**). Nonetheless, the industry faces difficulties because of high employee turnover rates brought on by unsatisfactory jobs, low compensation, few opportunities for professional growth, and a poor work-life balance. As a result, the industry needs more efficient talent management strategies to boost employee engagement, retention, and productivity (**Pizam and Shani, 2009; Ruel, 2018; Jooss, 2018**).

Particularly encouraging and dangerous for the industry in this sense is the advancement of machine/artificial intelligence (AI) capabilities (**Ivanov and Webster, 2017; OECD, 2018; Cain et al., 2019; Prentice et al., 2020**).

**According to de Leede (2016) and Ivanov and Webster (2019),** artificial intelligence (AI) enables people to work more intelligently, which improves business outcomes. However, it also necessitates the development of new competencies and capabilities, including technical expertise, social and emotional intelligence, and creative skills. However, AI has the potential to displace human talent with technology in some situations, which would force the hospitality and tourism industries to rethink their organizational structures and operational procedures **(Ivanov and Webster, 2017; OECD, 2018; Cain et al., 2019; Prentice et al., 2020)**

**AIM, OBJECTIVE AND RESEARCH QUESTIONS**

This research aims to explore and demonstrate the transformative potential of data analytics tools, including Python, SQL and using Business Intelligence tool Power BI, within the hospitality industry. By examining current data utilization practices, evaluating operational efficiency enhancements, and analysing the integration of data analytics into decision-making processes, this study seeks to provide a comprehensive understanding of how data-driven insights can revolutionize strategic decision-making in hospitality enterprises.

**OBJECTIVE:-**

1. **Examine the Current Data Utilization Practices in the Hospitality Industry:**
2. Conduct a thorough analysis of how hospitality organizations currently employ data analytics tools in decision-making processes.
3. Identify opportunities and challenges associated with utilizing data to inform decisions in the hospitality sector.
4. **Evaluate the Impact of Data Analytics on Operational Efficiency:**
5. Evaluate and understand the hospitality data by using Python and SQL for performing data analysis and use the business intelligence tool Power BI for creating and optimizing operational procedures within the hospitality industry.
6. Quantify the increase in productivity resulting from the adoption of data analytics tools.
7. **Examine Data Analytics-Enhanced Decision-Making Processes:**
8. Analyse case studies showcasing instances where data analytics tools have been instrumental in making strategic decisions.
9. Assess the ways in which these tools enhance forecasting, resource allocation, and overall decision quality.
10. **Conduct a Platform Comparison Analysis for Data Analytics:**
11. Compare and contrast the advantages and limitations of Power BI, Python, and SQL in the context of data analytics for the hospitality industry.
12. Provide recommendations on selecting the most suitable tools for different analytical scenarios.
13. **Explore Integration of Data Analytics into Business Strategy:**
14. Investigate how hospitality enterprises integrate insights from data analytics into their strategic planning processes.
15. Identify challenges and benefits associated with aligning data analytics with organizational objectives.

**Research Questions:**

1. To what extent do hospitality companies currently utilize data analytics tools in decision-making processes?
2. How do SQL, Python, Excel, and Power BI contribute to enhancing operational efficiency within the hospitality industry?
3. Which critical decision-making procedures in the hospitality sector stand to benefit most from the application of data analytics?
4. Which data analytics platform is most suitable for specific analytical requirements within the hospitality industry?
5. How can hospitality enterprises effectively integrate data analytics into their broader strategic planning procedures?

By addressing these aims, objectives, and research questions, this study aims to illuminate the potential of data analytics in transforming decision-making processes within the hospitality domain.

**METHODOLOGY AND IMPLEMENTATION**

In the implementation phase of this research, a multi-faceted approach was adopted to derive actionable insights from the dataset pertaining to the hospitality industry. Beginning with Python, a comprehensive Exploratory Data Analysis (EDA) was conducted, facilitating an in-depth exploration of the dataset's structure, patterns, and potential anomalies. Leveraging Python's extensive libraries for data analysis and visualization, this phase laid the groundwork for subsequent analytical endeavours. Following the EDA phase, the dataset was seamlessly integrated into the MSSQL server environment. This step ensured centralized access to the data, enabling efficient data management and scalability for future analyses. Utilizing SQL queries, a series of data exploration exercises were undertaken, allowing for targeted investigations into specific aspects of the dataset relevant to the hospitality domain. Transitioning from SQL-based exploration, the dataset was further processed and refined for visualization and reporting in Power BI. Leveraging the capabilities of Power BI, the imported Excel (CSV) datasets were meticulously structured to facilitate the creation of a dynamic and interactive dashboard. This dashboard serves as a focal point for encapsulating key insights and trends gleaned from the data, tailored specifically to meet the needs of stakeholders within the hospitality industry.

Central to the design of the Power BI dashboard is the strategic alignment of visual elements to convey meaningful insights effectively. Through the judicious selection of visualization types such as bar charts, line graphs, and heatmaps, complex data relationships are distilled into intuitive representations. These visualizations are thoughtfully organized within the dashboard layout, ensuring coherence and ease of interpretation for end-users. Moreover, the Power BI dashboard is designed to empower stakeholders with the ability to interactively explore the dataset. Through user-friendly features such as slicers, filters, and drill-down capabilities, individuals can dynamically manipulate the data presentation to uncover deeper insights and investigate specific trends of interest. In addition to static visualizations, the dashboard incorporates dynamic elements such as KPI indicators and trend analysis tools. These components provide stakeholders with real-time performance metrics and facilitate longitudinal comparisons, enabling proactive decision-making within the hospitality industry. Furthermore, the implementation phase encompasses a rigorous validation process to ensure the accuracy and reliability of the insights generated. Through iterative testing and validation against known benchmarks or industry standards, the integrity of the analytical outputs is upheld, instilling confidence in the findings presented.

Overall, the implementation of this data analysis pipeline represents a holistic approach to deriving actionable insights for the hospitality industry. By leveraging the synergistic capabilities of Python, MSSQL, and Power BI, a comprehensive understanding of the dataset is achieved, empowering stakeholders with the knowledge needed to make informed decisions and drive meaningful outcomes in the dynamic landscape of hospitality.

**PROBLEM STATEMENT**

In the fiercely competitive landscape of luxury/business hotels in India, AtliQ Grands, a stalwart in the hospitality industry with a 20-year legacy, faces a formidable challenge. Despite its longstanding presence and esteemed reputation, AtliQ Grands is witnessing a concerning decline in both market share and revenue. This decline can be attributed to a combination of strategic manoeuvres by competitors and shortcomings in decision-making within the management echelons.

Recognizing the urgency of the situation, the managing director of AtliQ Grands has embarked on a strategic initiative: the integration of Business and Data Intelligence into the company's operational framework. The objective is clear—to leverage data-driven insights as a catalyst for reclaiming lost market share and optimizing revenue streams. However, the absence of an in-house data analytics team presents a significant obstacle to this endeavour.

To bridge this gap, AtliQ Grands' revenue management team has opted to enlist the expertise of a third-party service provider. This decision underscores the critical importance of data-driven insights in informing strategic decision-making and propelling AtliQ Grands back to the forefront of the luxury/business hotels segment.

***Dataset Overview :-***

AtliQ Grands, a revered name in India's hospitality landscape with a 20-year legacy, faces a formidable challenge amidst the fiercely competitive luxury/business hotels segment. Despite its esteemed reputation, AtliQ Grands is witnessing a concerning decline in market share and revenue, attributed to both competitive manoeuvres and internal decision-making shortcomings. In response, the managing director has initiated a transformative journey, aiming to leverage Business and Data Intelligence as a strategic imperative to reclaim dominance.

***Challenges and Strategic Imperatives:***

The decline in market share and revenue necessitates a strategic overhaul, emphasizing data-driven insights to inform decision-making. However, the absence of an in-house data analytics team presents a significant obstacle. To bridge this gap, AtliQ Grands' revenue management team has decided to engage a third-party service provider. This underscores the critical importance of data-driven insights in guiding strategic decisions and propelling AtliQ Grands back to the forefront of the luxury/business hotels segment.

***Dataset Description:***

The dataset comprises meticulously curated historical records spanning AtliQ Grands' portfolio of five-star hotels across India. It encompasses a comprehensive array of operational metrics, guest demographics, booking patterns, revenue streams, and market dynamics. The dataset is organized into five CSV files, each containing essential information crucial for deriving actionable insights:

* ***dim\_date:*** Contains dates in May, June, and July, along with week numbers and day types (Weekend/Weekday).
* ***dim\_hotels:*** Provides unique IDs, names, categories (Luxury/Business), and locations of AtliQ Grands' hotels.
* ***dim\_rooms:*** Specifies room types and their corresponding classes (Standard, Elite, Premium, Presidential).
* ***fact\_aggregated\_bookings:*** Presents aggregated booking data including hotel IDs, check-in dates, room categories, successful bookings, and capacities.
* ***fact\_bookings:*** Includes individual booking records with details such as booking IDs, hotel IDs, booking dates, check-in/check-out dates, guest counts, room categories, booking platforms, ratings, booking statuses, and revenue information.

The integration of Business and Data Intelligence represents a pivotal strategic initiative for AtliQ Grands in its quest to reclaim dominance in India's luxury/business hotels segment. . By harnessing the power of data-driven insights, AtliQ Grands endeavours to not only address current challenges but also position itself as a leader in the ever-evolving hospitality landscape.The primary goal of this dataset is to equip AtliQ Grands with actionable insights derived from historical data.

By leveraging Business and Data Intelligence, AtliQ Grands aims to revitalize market share and revenue growth in India's luxury/business hotels sector. Informed decision-making based on empirical evidence will enable AtliQ Grands to adapt to market dynamics, enhance operational efficiency, optimize revenue streams, and deliver unparalleled guest experiences. Through proactive decision-making and continuous innovation, AtliQ Grands is poised to usher in a new era of success and excellence. Here you can get familiarized with the content more properly (the respective stages of the project are indicated in parentheses) :

* Exploratory analysis on the dataset using Python.
* Data Visualization and analysis using Python.

**EXPLORATORY DATA ANALYSIS USING PYTHON**

**DATA ANALYSIS &VISUALISATION USING PYTHON**

***Exploratory Data Analysis using Python:-***

Exploratory Data Analysis (EDA) is a critical first step in the data analysis process. Its primary objective is to summarize the main characteristics of the data, gain insights into the underlying structure, detect anomalies, and formulate hypotheses for further investigation. Here's a brief overview of some common methods used in EDA:

1. ***Summary of the Statistics:*** This involves computing descriptive statistics such as mean, median, mode, standard deviation, minimum, maximum, and quantiles for each variable in the dataset. These statistics provide a quick snapshot of the data distribution.
2. ***Data Visualization:*** Visualization techniques such as histograms, box plots, scatter plots, and bar plots help to visualize the distribution, central tendency, variability, and relationships between variables in the dataset. Visualizations are essential for identifying patterns, trends, outliers, and potential relationships within the data.
3. ***Correlation Analysis:*** Correlation analysis measures the strength and direction of the linear relationship between pairs of variables. Pearson correlation coefficient is commonly used for continuous variables, while Spearman or Kendall rank correlation coefficients are used for ordinal or non-normally distributed data.
4. ***Outlier Detection:*** Outliers are data points that significantly deviate from the rest of the data. Techniques such as visualization, z-score, Tukey's method, and clustering-based approaches can help identify outliers, which may require further investigation to determine if they are genuine data points or errors.
5. **Missing Values Handling:** Missing data can significantly affect the analysis and interpretation of results. EDA involves identifying missing values and deciding on appropriate strategies for handling them, such as imputation, deletion, or modelling techniques.
6. **Feature Engineering:** EDA can also involve creating new features or transforming existing ones to better represent the underlying patterns in the data. Techniques such as scaling, encoding categorical variables, and creating interaction terms can enhance the predictive power of the models.
7. **Dimensionality Reduction:** In cases where the dataset has a large number of features, dimensionality reduction techniques such as principal component analysis (PCA) or t-distributed stochastic neighbour embedding (t-SNE) can help visualize high-dimensional data and identify the most important features.
8. **Clustering Analysis:** Clustering techniques such as K-means clustering or hierarchical clustering can be applied to identify natural groupings or clusters within the data. This can provide insights into the underlying structure of the dataset and aid in segmentation or pattern recognition.

Overall, exploratory data analysis plays a crucial role in understanding the characteristics of the data, guiding subsequent modelling decisions, and generating actionable insights for decision-making. It's an iterative process that often leads to further questions and refinements as more insights are uncovered during the analysis.

Data analysis and visualization are indispensable components of modern data science, empowering analysts and researchers to extract insights, identify patterns, and communicate findings effectively. In the realm of Python, an extensive ecosystem of libraries and tools exists to facilitate these tasks, offering versatile solutions for a wide range of data analysis needs. This dissertation provides a comprehensive exploration of key data analysis and visualization techniques in Python, focusing on prominent methods such as box plots, pie charts, histograms, bar plots, and leveraging the Seaborn library for enhanced visualization capabilities. The most popular programming language for Data Science or ML or AI is using either Python (or) R, both are highly user friendly, flexible, open source and useful for data analysis. Python is the most powerful language in terms of implementing it in ML due to its flexibility. In this project I have utilized Python programming with some of the common python libraries for Data exploration, Data Analysis and Visualization. Some of the common python libraries which were used in this project are Pandas, NumPy, Seaborn and Matplotlib. There were lot of development Environments available in the market out of which I choose google collab Integrated development environment. Because one can share code notebooks easily and there is no need to install any modules to run any code, modules come preinstalled within Google Collab.

**Google Collab Environment:**

Data analysis and visualization using Python on Google Collab offer a powerful and accessible approach to exploring datasets, gaining insights, and communicating findings. By leveraging the capabilities of Python libraries within the collaborative environment of Google Collab, analysts can efficiently conduct EDA, create informative visualizations, and drive data-driven decision-making processes. Google Collab provides a convenient platform for data analysis and coding in Python. Key advantages include:

* Collaboration: Share code notebooks easily with collaborators.
* No Installation Required: Pre-installed modules eliminate the need for manual installation.
* Performance: High-performance computing capabilities.
* Integration with GitHub: Easily save and publish code notebooks to GitHub repositories.

**Python Libraries:**

Python offers a rich ecosystem of libraries and tools for data analysis and visualization, making it a preferred choice for researchers, analysts, and practitioners across various domains. By leveraging techniques such as box plots, pie charts, histograms, bar plots, and Seaborn visualizations, analysts can uncover insights, communicate findings, and drive data-driven decision-making processes effectively. Through practical examples and demonstrations, this dissertation aims to equip readers with the knowledge and skills necessary to harness the power of Python for data analysis and visualization tasks.

Python offers a rich ecosystem of libraries for data analysis and visualization. Commonly used libraries in this project include are:

* **Pandas:-** For data manipulation and analysis.
* **NumPy:-** For numerical computing and array operations.
* **Seaborn:**- For statistical data visualization.
* **Matplotlib:**- For creating static, interactive, and publication-quality visualizations.

**Pandas:-** It is a powerful library for data manipulation and analysis. It provides data structures and functions to efficiently manipulate structured data, such as tabular data (e.g., CSV files, Excel sheets). Key Features: Data Frame and Series objects for storing and manipulating data, powerful methods for data cleaning, reshaping, merging, and grouping.

**NumPy:-** It is a fundamental python library used for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently.

Key Features: Multi-dimensional array objects (ndarray), functions for array manipulation, mathematical operations, linear algebra operations, and random number generation.

**Seaborn:-** It is a statistical data visualization library based on Matplotlib. It provides a high-level interface for creating informative and attractive statistical graphics.

Key Features:- Simplified syntax for creating complex plots, support for various plot types (e.g., scatter plots, bar plots, box plots), built-in themes and colour palettes, integration with Pandas Data Frames.

**Matplotlib:-** It is a comprehensive library for creating static, interactive, and animated visualizations in Python. It serves as the foundation for many other plotting libraries in Python.

Key Features: Support for various plot types (e.g., line plots, scatter plots, histograms), fine-grained control over plot elements, ability to create publication-quality plots, support for multiple backends for rendering plots in different environments.

A screenshot of a computer code

Description automatically generated

**Loading and Importing Datasets:-**

A screenshot of a computer code

Description automatically generated

data is the variable which used to store, save and load the dataset “***dim\_date”***

* ***dim\_date:*** Contains dates in May, June, and July, along with week numbers and day types (Weekend/Weekday).

data\_hotels is the variable which used to store, save and load the dataset “***dim\_hotels”***

* ***dim\_hotels:*** Provides unique IDs, names, categories (Luxury/Business), and locations of AtliQ Grands' hotels.

data\_rooms is the variable which used to store, save and load the dataset “***dim\_rooms”***

* ***dim\_rooms:*** Specifies room types and their corresponding classes (Standard, Elite, Premium, Presidential).

data\_agg is the variable which used to store, save and load the dataset “***fact\_aggregated\_bookings”***

* ***fact\_aggregated\_bookings:*** Presents aggregated booking data including hotel IDs, check-in dates, room categories, successful bookings, and capacities. data\_fact is the variable which used to store, save and load the dataset “fact\_aggregated\_bookings”
* ***fact\_bookings:*** Includes individual booking records with details such as booking IDs, hotel IDs, booking dates, check-in/check-out dates, guest counts, room categories, booking platforms, ratings, booking statuses, and revenue information.

**data.head(5):** This function returns the first 5 rows of the Data frame data. It's useful for quickly inspecting the structure and contents of your DataFrame.

**data.tail(5):** Conversely, this function returns the last 5 rows of the Data frame data. It's helpful when you want to see the tail end of your data.

A screenshot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated**Checking Null values:-**

A screenshot of a computer

Description automatically generated

We could see there missing in the column “ratings\_given” for the dataset ***fact\_bookings.*** This has been handled using one of the data cleaning fill(0)na. The purpose of fillna(0) function is used to replace any missing (NaN) values in a DataFrame or Series with a specified value, in this case, 0. By applying fillna(0) to the 'ratings\_given' column of the DataFrame data\_fact, any NaN values in that column will be replaced with 0. This ensures that the column contains numeric values throughout, which can be important for certain calculations or analyses where missing values could cause errors or inconsistencies.

A screenshot of a computer

Description automatically generated

Now you can see there are no missing values present in the data\_fact dataset.

The provided code is changing the data type of specific columns in different Dataframes from date to datetime format.

A screenshot of a computer code

Description automatically generated

Performing Data Analysis on Hotel Revenue:-

In order to perform the analysis on hotel information to crack with some facts or statistics about these hotels we need to combine the two DataFrames (data\_hotels and data\_fact) based on a common column property\_id. The how='left' parameter specifies a left join, meaning all the rows from the left DataFrame (data\_hotels) will be retained, and matching rows from the right DataFrame (data\_fact) will be joined. If there are no matching rows in the right DataFrame, NaN values will be filled. After merging, the .head() method is used to display the first few rows of the resulting DataFrame rev.

A screenshot of a computer

Description automatically generated

**Hotel Wise Revenue:-**

A screenshot of a computer

Description automatically generated

A bar graph with different colored bars

Description automatically generatedA screen shot of a computer code

Description automatically generated

A computer code with text

Description automatically generated with medium confidence

A bar graph with different colored bars

Description automatically generated

A chart of different colors

Description automatically generatedA computer code with text

Description automatically generated

A screenshot of a computer

Description automatically generated

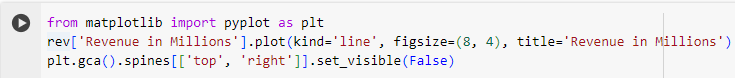
A computer code with text

Description automatically generatedA diagram of different colored shapes

Description automatically generated

A graph of different colored bars

Description automatically generated

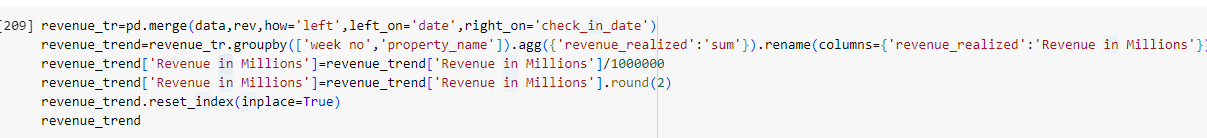


A graph with a line

Description automatically generated

A screenshot of a computer

Description automatically generated



A graph with lines and text

Description automatically generated

A screenshot of a computer code

Description automatically generated

A diagram of different colored shapes

Description automatically generated with medium confidence

A close-up of a computer screen

Description automatically generated

A graph of a chart

Description automatically generated with medium confidence

A screen shot of a graph

Description automatically generated

A computer code with text

Description automatically generated with medium confidence

A graph with orange and blue lines

Description automatically generated

**Booking Analysis-**

A screenshot of a computer

Description automatically generated

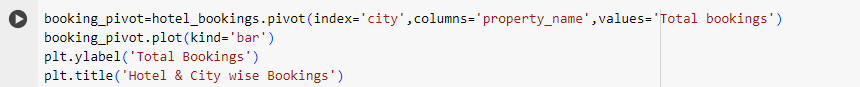
A screen shot of a graph

Description automatically generatedA graph with text and numbers

Description automatically generatedA screen shot of a graph

Description automatically generated A screen shot of a graph

Description automatically generated



A graph of different colored bars

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A graph with numbers and dots

Description automatically generatedA screen shot of a computer screen

Description automatically generated

A screen shot of a graph

Description automatically generatedA screenshot of a computer screen

Description automatically generatedA screenshot of a graph

Description automatically generatedA graph with text and numbers

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A white background with red text

Description automatically generated

A graph with text on it

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generatedA graph with text and numbers

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph with lines and text

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A graph with numbers and dots

Description automatically generated



A pie chart with numbers and a few different colored circles

Description automatically generated with medium confidence

A screenshot of a computer screen

Description automatically generated

A graph of hotels and hotels

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screen shot of a computer

Description automatically generated

A screen shot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generatedA graph with lines and text

Description automatically generated

A graph with lines and numbers

Description automatically generatedA graph with lines and text

Description automatically generatedA screen shot of a graph

Description automatically generated A screenshot of a graph

Description automatically generated

A graph with lines and text

Description automatically generatedA screenshot of a computer screen

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph of a bar graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generatedA screen shot of a graph

Description automatically generated

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer screen

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

A graph with blue and white text

Description automatically generated

**Exploratory Data Analysis using SQL:-**

SQL, or Structured Query Language, is a powerful tool in the realm of data analysis. It serves as a standardized language for managing and manipulating relational databases, making it indispensable for anyone working with structured data. In this brief exploration, we'll delve into the various ways SQL facilitates data analysis, from data retrieval and manipulation to aggregation and reporting.

**Data Retrieval:** One of the fundamental capabilities of SQL is its ability to retrieve data from databases. Using the SELECT statement, analysts can specify the columns they want to retrieve and apply filters to narrow down the results. This feature enables analysts to extract relevant data subsets for further analysis.

**Data Manipulation:** SQL provides a robust set of tools for manipulating data within a database. Analysts can use SQL statements like INSERT, UPDATE, and DELETE to add, modify, or remove records from tables. For instance, an analyst might use the UPDATE statement to adjust product prices based on market trends or customer demand. This capability allows for the transformation of raw data into formats suitable for analysis, ensuring data integrity and consistency throughout the process.

**Data Aggregation:** Aggregating data is a crucial aspect of data analysis, and SQL offers several functions to facilitate this process. Functions like COUNT, SUM, AVG, MIN, and MAX allow analysts to summarize data and derive meaningful insights. Aggregating data enables analysts to identify trends, patterns, and anomalies within large datasets, providing valuable insights for decision-making.

**Data Joins:** Relational databases often consist of multiple tables linked by common keys, and SQL excels at performing joins to combine data from these tables. Joins allow analysts to create comprehensive datasets by merging information from related tables. For instance, to analyze sales data along with customer demographics, one might perform a join between the Sales and Customers tables using a common key like CustomerID. SQL supports various types of joins, including INNER JOIN, LEFT JOIN, and RIGHT JOIN, giving analysts flexibility in how they merge data from different sources.

**Data Filtering and Sorting:** SQL enables analysts to filter and sort data based on specific criteria, allowing for targeted analysis. The WHERE clause allows analysts to specify conditions that records must meet to be included in the result set. Additionally, the ORDER BY clause enables sorting the results based on one or more columns, either in ascending or descending order. Filtering and sorting capabilities help analysts focus on relevant data and identify trends more effectively.

**Data Grouping and Summarization:** Grouping and summarizing data are essential tasks in data analysis, and SQL provides powerful features to accomplish these tasks. The GROUP BY clause allows analysts to group rows based on one or more columns, while aggregate functions can then be applied to each group to calculate summary statistics. For instance, to calculate the total sales revenue for each product category, one could use the following query: SELECT Category, SUM(Revenue) FROM Sales GROUP BY Category;. Grouping and summarization enable analysts to gain insights into overall trends and patterns within the data.

**Data Validation and Quality Assurance:** Ensuring data quality is crucial for meaningful analysis, and SQL offers mechanisms for data validation and quality assurance. Constraints such as UNIQUE, NOT NULL, and CHECK can be enforced at the database level to prevent invalid or inconsistent data from being entered. Additionally, SQL provides tools for data cleansing and transformation, allowing analysts to identify and correct errors or inconsistencies in the data. By maintaining data integrity and quality, analysts can trust the results of their analyses and make informed decisions based on reliable information.

**Data Reporting and Visualization:** While SQL itself is primarily a query language for interacting with databases, it plays a crucial role in the data reporting and visualization process. Analysts can use SQL to retrieve the necessary data from databases and then feed it into reporting and visualization tools for further analysis and presentation. SQL queries can also be embedded within reporting applications or used to generate data extracts for use in other analytical tools. By leveraging SQL in conjunction with reporting and visualization tools, analysts can create informative reports and dashboards that communicate insights effectively to stakeholders.

**Scalability and Performance:** SQL databases are designed to handle large volumes of data efficiently, making them well-suited for scalable data analysis. SQL engines employ optimization techniques such as indexing, query caching, and query optimization to enhance performance and ensure timely responses to queries, even with massive datasets. This scalability and performance make SQL a preferred choice for organizations dealing with substantial amounts of data, enabling them to perform complex analyses without sacrificing speed or efficiency.

**Purpose and Reason for choosing MSSQL Database:-**

Choosing Microsoft SQL Server (MSSQL) for data exploration and analysis can be driven by several factors, including its robust features, scalability, integration capabilities, and ease of use for those already familiar with Microsoft technologies. Here are some reasons why MSSQL might be chosen for data exploration and analysis:

**Integration with Microsoft Ecosystem:** MSSQL offers seamless integration with other Microsoft tools such as Excel, Power BI, and Azure services, facilitating data analysis and reporting workflows.

**Scalability and Performance:** MSSQL is known for its scalability, allowing organizations to handle large volumes of data efficiently. It offers features like partitioning, indexing, and query optimization to ensure high performance even with complex analytical queries.

**Advanced Analytics:** MSSQL provides built-in support for advanced analytics through features like SQL Server Analysis Services (SSAS), which allows for multidimensional data analysis, and SQL Server Integration Services (SSIS), which enables data integration and transformation tasks.

**Security:** MSSQL offers robust security features to protect sensitive data, including encryption, access controls, and auditing capabilities. This is essential for organizations dealing with confidential or regulated data.

**Reliability and Support:** As a widely used relational database management system (RDBMS), MSSQL benefits from a large community of users and extensive documentation. Additionally, Microsoft provides comprehensive support and regular updates to ensure reliability and address any issues promptly.

**Advanced Querying Capabilities:** MSSQL supports complex SQL queries, stored procedures, and user-defined functions, making it suitable for sophisticated data analysis tasks. It also offers features like Common Table Expressions (CTEs) and window functions for advanced data manipulation.

**Data Visualization:** While MSSQL itself is primarily a database management system, it integrates seamlessly with data visualization tools like Power BI, allowing users to create interactive dashboards and reports to visualize insights derived from the data.

**Hybrid and Cloud Options:** With the increasing adoption of cloud technologies, MSSQL offers both on-premises and cloud-based deployment options. Organizations can choose to deploy MSSQL on Microsoft Azure for scalability, flexibility, and cost-efficiency.

Overall, the choice of MSSQL for data exploration and analysis often boils down to factors such as compatibility with existing systems, performance requirements, security needs, and the availability of skilled personnel.

**Importing Data into "Hospitality Database" on MSSQL SERVER:**

The journey commences with the creation of a dedicated database, aptly named "Hospitality Database," within the Microsoft SQL Server environment. This serves as the repository for our imported datasets, which were initially stored in CSV format. Leveraging SQL Server Management Studio (SSMS), the Import/Export Wizard facilitates seamless migration of data from CSV files into corresponding SQL database tables.

The imported datasets encompass diverse facets of hospitality operations, ranging from guest demographics to reservation details and service feedback. Each dataset is represented as a relational table within the "Hospitality Database," meticulously structured to maintain data integrity and facilitate efficient querying.

Armed with SQL prowess, we embark on an analytical journey to glean insights and unravel key metrics embedded within the datasets. Through a series of basic to intermediate level SQL queries, we unearth valuable nuggets of information across various dimensions:

**Guest Demographics:** By querying the "Guests" table, we discern the demographic composition of clientele, including geographical distribution, age demographics, and customer segmentation based on visit frequency.

**Reservation Dynamics:** Delving into the "Reservations" table unveils critical reservation metrics such as booking trends over time, average lead time, booking channels, and peak occupancy periods.

**Revenue Analysis:** Leveraging transactional data from the "Transactions" table, we conduct revenue analysis to identify top-selling products/services, revenue contribution by customer segments, and trends in revenue generation over specified time intervals.

**Service Feedback and Ratings:** The "Feedback" table serves as a treasure trove of customer sentiment and satisfaction levels. Through sentiment analysis and aggregation of feedback scores, we gauge overall service performance and identify areas for improvement.

**Operational Efficiency:** Analyzing operational metrics such as table turnover rates, staff utilization, and service response times enables us to optimize resource allocation and streamline operational workflows for enhanced efficiency.

**Used the below SQL Queries for exploring the tables *dim\_date, dim\_hotels, dim\_rooms, fact\_aggregated\_bookings and fact\_bookings*** **using the MSSQL Server Database**

|  |
| --- |
| 1. **Viewing sample data:**   SELECT COUNT(\*) **AS total\_rows** FROM [Hospitality Database].[dbo].[dim\_date];   1. **Counting total rows:**   SELECT TOP 10 \*FROM [HospitalityDatabase].[dbo].[dim\_date];   1. **Checking for Duplicates**   SELECT **'Duplicates Exist' AS status WHERE EXISTS** (SELECT mmm\_yy, COUNT(\*) FROM [Hospitality Database].[dbo].[dim\_date] GROUP BY mmm\_yy HAVING COUNT(\*) > 1 );   1. **Finding the earliest and latest date:**   SELECT **MIN(date) AS earliest\_date, MAX(date) AS latest\_date** FROM [Hospitality Database].[dbo].[dim\_date]  **5. Finding the number of unique months and years:**  SELECT COUNT(DISTINCT mmm\_yy) AS unique\_month\_years FROM **[Hospitality Database].[dbo].[dim\_date]**  **6. Aggregating by week number:**  SELECT week\_no, **COUNT(\*)** AS num\_dates **FROM [Hospitality Database].[dbo].[dim\_date]** GROUP BY week\_no ORDER BY week\_no;  **7. Filtering by day type:**  SELECT day\_type, **COUNT(\*)** AS num\_days **FROM [Hospitality Database].[dbo].[dim\_date]** GROUP BY day\_type;  **8. Finding missing dates:**  WITH AllDates AS ( SELECT MIN(date) AS min\_date, MAX(date) AS max\_date FROM [Hospitality Database].[dbo].[dim\_date]) SELECT **DATEADD(day, number, min\_date)** AS missing\_date FROM AllDates  **JOIN** master..spt\_values ON DATEADD(day, number, min\_date) <= max\_date WHERE type = 'P' AND NOT EXISTS ( SELECT FROM [Hospitality Database].[dbo].[dim\_date] **WHERE dim\_date.date = DATEADD(day, number, min\_date));**  **9. Calculating average days between dates:**  SELECT AVG(DATEDIFF(day, date, LEAD(date, 1) OVER (ORDER BY date))) AS avg\_days\_between\_dates  FROM [Hospitality Database].[dbo].[dim\_date];  **10 . Identifying weekends:**  SELECT **COUNT(\*)** AS num\_weekends **FROM** [Hospitality Database].[dbo].[dim\_date] WHERE DAYNAME(date) **IN** ('Saturday', 'Sunday');  ***dim\_hotels:-***  1. **Basic Data Retrieval:** Retrieve all columns for all rows in the table:  SELECT \* **FROM** [Hospitality Database].[dbo].[dim\_hotels]  **2. Filtering Data:** Retrieve hotels located in a specific city:  SELECT \* **FROM** [Hospitality Database].[dbo].[dim\_hotels] WHERE city = ‘ ’;  **3. Aggregate Functions:** Count the number of hotels in each category:  SELECT category, COUNT(\*) AS total\_hotels **FROM** [Hospitality Database].[dbo].[dim\_hotels] GROUP BY category;  4. **Sorting Data:** Retrieve hotels sorted by property name in ascending order:  SELECT \* **FROM** [Hospitality Database].[dbo].[dim\_hotels] **ORDER BY** property\_name ASC;  **5. Joins:** If you have another table fact\_bookings containing booking information with a property\_id column, you can join it with dim\_hotels to get bookings along with hotel details:  SELECT b.\*, h.property\_name, h.city **FROM** fact\_bookings b **INNER JOIN** dim\_hotels h **ON** b.property\_id = h.property\_id;  **6. Subqueries:** Retrieve hotels that have more than 50 bookings:  SELECT \***FROM** [Hospitality Database].[dbo].[dim\_hotels] WHERE property\_id IN (SELECT property\_id FROM fact\_bookings **GROUP BY** property\_id HAVING COUNT(\*) > 50 );  **7. Conditional Logic:** Retrieve hotels with a category of 'Luxury' and located in a specific city:  SELECT **\*FROM** [Hospitality Database].[dbo].[dim\_hotels] WHERE category = 'Luxury' AND city = 'YourCity';  **8. Data Modification:** Update the category of a specific hotel:  UPDATE dim\_hotels SET category = 'NewCategory' WHERE property\_id = 'YourPropertyID';  **Data Removal:** Delete hotels that are not in use (assuming no bookings are present):  DELETE FROM dim\_hotels WHERE property\_id NOT IN (SELECT DISTINCT property\_id FROM fact\_bookings);  ***fact\_bookings***  **1. Total number of bookings:**  SELECT COUNT(\*) AS total\_bookings FROM [Hospitality Database].[dbo].[fact\_bookings];  **2. Total revenue generated:**  SELECT SUM (revenue\_generated) AS total\_revenue FROM [Hospitality Database].[dbo].[fact\_bookings];  **3. Average revenue per booking:**  SELECT AVG(revenue\_generated) AS avg\_revenue\_per\_booking  **FROM** [Hospitality Database].[dbo].[fact\_bookings];  **4. Total revenue realized:**  SELECT SUM( revenue\_realized) **AS** total\_revenue\_realized FROM [Hospitality Database].[dbo].[fact\_bookings];  **5. Average ratings given:**  SELECT AVG(ratings\_given) AS avg\_ratings\_given **FROM** [Hospitality Database].[dbo].[fact\_bookings];  **6. Number of bookings per platform:**  SELECT booking\_platform, COUNT(\*) AS total\_bookings  FROM [Hospitality Database].[dbo].[fact\_bookings] **GROUP BY** booking\_platform;  **7. Revenue generated per room category:**  **SELECT** room\_category, SUM(revenue\_generated) **AS** total\_revenue  **FROM** [Hospitality Database].[dbo].[fact\_bookings] **GROUP BY** room\_category;  **8. Number of bookings and revenue generated per month:**  **SELECT** MONTH(booking\_date) AS booking\_month, COUNT(\*) **AS** total\_bookings,  **SUM**(revenue\_generated) **AS** total\_revenue **FROM** [Hospitality Database].[dbo].[fact\_bookings] **GROUP BY MONTH**(booking\_date);  **9. Number of bookings and revenue generated per property:**  **SELECT** property\_id, COUNT(\*) AS total\_bookings, SUM(revenue\_generated) AS total\_revenue **FROM** [Hospitality Database].[dbo].[fact\_bookings] **GROUP BY** property\_id;  **10. Number of bookings and revenue realized per property:**  **SELECT** property\_id, COUNT(\*) AS total\_bookings, SUM(revenue\_realized) AS total\_revenue\_realized  **FROM** [Hospitality Database].[dbo].[fact\_bookings] **GROUP BY** property\_id;  ***fact\_aggeregated\_bookings***  **1. Count the number of records in the table:** SELECT **COUNT(\*)** AS total\_records **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **2. Display the first 10 records in the table:**  SELECT **TOP** 10 \***FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **3. Find the distinct room categories available:**  SELECT **DISTINCT** room\_category **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **4. Calculate the total number of successful bookings:**  SELECT **SUM** (successful\_bookings) **AS** total\_successful\_bookings **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **5. Find the total capacity of all properties:**  SELECT **SUM(capacity)** AS total\_capacity **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **6. List properties with the highest and lowest capacity:**  SELECT **TOP** property\_id, capacity **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **ORDER BY** capacity DESC; -- Highest capacity SELECT **TOP** 1 property\_id, capacity **FROM** [Hospitality  Database].[dbo].[fact\_aggregated\_bookings] ORDER BY capacity ASC; -- Lowest capacity  **7. Calculate the average capacity per room category:**  SELECT room\_category, AVG(capacity) **AS** average\_capacity  **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings] **GROUP BY** room\_category;  8**. Find the date with the highest number of successful bookings:**  **SELECT** check\_in\_date, MAX(successful\_bookings) **AS** max\_successful\_bookings  **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings];  **9. Find the date with the lowest number of successful bookings:**  **SELECT** check\_in\_date, MIN(successful\_bookings) **AS** min\_successful\_bookings **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings]  **11. Calculate the success rate for each room category:**  **SELECT** room\_category, SUM(successful\_bookings) \* 100.0 / SUM(capacity) AS success\_rate\_percentage  **FROM** [Hospitality Database].[dbo].[fact\_aggregated\_bookings] **GROUP BY** room\_category; |

**INTRODUCTION TO POWERBI:**

Power BI is a cloud-based business analytics service that enables users to visualize and share insights from their data in a unified platform. With intuitive dashboards, interactive reports, and rich visualizations, Power BI empowers users to make data-driven decisions, uncover patterns, and identify trends.

**POWER QUERY**

Power Query is a data connectivity and preparation tool that allows users to easily discover, connect, and transform data from various sources into a format suitable for analysis. Key features of Power Query include data cleansing, transformation, and data shaping capabilities. By providing a seamless interface for data manipulation, Power Query simplifies the process of preparing data for analysis, saving time and effort for analysts.

**DAX (Data Analysis Expressions)**

Dax serves as a powerful formula language in PowerBI, allowing users to create custom calculations, measures, and calculated columns to derive insights from the data, thus enhancing the analytical capabilities of the platform.

**METRICS**

Metrics play a crucial role in measuring performance and tracking key performance indicators (KPIs) within an organization. Power BI enables users to create and visualize metrics using a wide range of data sources, including databases, spreadsheets, and cloud services. With customizable dashboards and real-time updates, users can monitor business performance, identify areas for improvement, and make data-driven decisions to drive success.

**BEST PRACTICES FOR POWEBI REPORTING:**

To maximize the effectiveness of Power BI reporting, it's important to follow best practices for data visualization, report design, and data governance. Some key best practices include:

* Designing intuitive and user-friendly dashboards and reports.
* Choosing appropriate visualizations to effectively communicate insights.
* Ensuring data accuracy and consistency through data governance policies.
* Collaborating and sharing insights with stakeholders through Power BI sharing and collaboration features.

**IMPORTING DATA TO POWERBI**

PowerBI supports a wide range of data formats, including CSV, Excel, XML, JSON, SQL databases (like MySQL, SQL Server, PostgreSQL), SharePoint lists, and cloud services like Azure, Google Analytics, Salesforce, allowing users to work with data in their preferred format. Importing datasets into PowerBI is a seamless process, allowing users to connect to various data sources such as Excel files, databases, cloud services, and more, enabling a comprehensive analysis of diverse datasets.

**DATA ANALYSIS WITH POWERBI:**

Power BI offers powerful data analysis capabilities that allow users to explore, analyze, and visualize data in meaningful ways. With features such as interactive visualizations, advanced analytics, and natural language querying, Power BI empowers users to uncover insights, gain deeper understanding of their data, and communicate findings effectively. Whether it's forecasting sales trends, identifying outliers, or segmenting customer data, Power BI provides the tools needed to drive actionable insights and drive business growth.

* PowerBI provides interactive visualizations such as charts, graphs, maps, and tables, allowing users to explore data dynamically.
* Its drag-and-drop interface makes it easy to create compelling visualizations without the need for complex coding.
* Users can drill down into data, apply filters, and create custom views to gain deeper insights into their data.

**Benefits of PowerBI for Data Analysis:**

* PowerBI accelerates the data analysis process by providing real-time insights into business performance.
* It empowers users to make informed decisions by presenting data in a visually appealing and understandable manner.
* The ability to create interactive dashboards and reports enables stakeholders to monitor key metrics and KPIs effortlessly.

**Benefits of using PowerBI in the Hospitality Sector:**

In the hospitality sector, PowerBI proves to be invaluable for data analysis by allowing businesses to analyze guest demographics, booking patterns, revenue trends, customer feedback, and operational metrics to optimize various aspects of their operations. It offers robust sharing and collaboration features, allowing users to publish reports and dashboards to the Power BI service or share them directly with the consumers and stakeholders. This facilitates seamless collaboration and enables teams to work together on analyzing and interpreting data in real-time.

* In the hospitality sector, PowerBI can revolutionize data analysis by aggregating data from various sources such as reservations, bookings, customer feedback, and operational metrics.
* It enables hoteliers and restaurant managers to track occupancy rates, revenue per available room (RevPAR), average daily rate (ADR), and other key performance indicators in real-time.
* PowerBI's predictive analytics capabilities can help forecast demand, optimize pricing strategies, and improve resource allocation.
* It allows businesses to segment customers based on preferences, demographics, and booking patterns, leading to targeted marketing campaigns and personalized guest experiences.
* PowerBI's ability to integrate with other Microsoft tools such as Excel, SharePoint, and Dynamics 365 further enhances its utility in the hospitality sector.
* PowerBI emerges as a transformative tool for data analysis in the hospitality sector, empowering businesses to unlock the full potential of their data and gain a competitive edge in today's dynamic market landscape.
* PowerBI's scalable architecture makes it suitable for businesses of all sizes, from small boutique hotels to large hotel chains, allowing them to harness the power of data analytics to drive growth and innovation in the ever-evolving hospitality industry.
* PowerBI's intuitive interface and interactive visualizations make data analysis a breeze, empowering users to explore trends, patterns, and correlations within the data quickly and efficiently.
* With its robust reporting capabilities, PowerBI enables users to create insightful dashboards and reports that provide actionable insights, facilitating data-driven decision-making processes.
* By leveraging PowerBI's advanced analytics features, hospitality businesses can identify emerging trends, forecast demand, personalize guest experiences, and optimize resource allocation to enhance customer satisfaction and drive profitability.
* PowerBI's integration with other Microsoft products such as Azure, Dynamics 365, and Office 365 further extends its capabilities, enabling seamless data integration and collaboration across different departments within hospitality organizations.
* The ability to schedule data refreshes and automate report generation in PowerBI ensures that hospitality businesses have access to real-time insights, enabling them to make timely decisions and stay ahead of the competition.
* With its mobile compatibility, PowerBI empowers hospitality professionals to access critical insights on-the-go, facilitating informed decision-making even when they are away from their desks.

A close-up of a graph

Description automatically generated**Created a Dashboard using POWERBI for Hospitality Sector:-**

**How much do hospitality companies currently use data analytics tools when making decisions?**

Hospitality companies are increasingly leveraging data analytics tools like Power BI for decision-making processes. The ability to build custom visuals, analyse trends, and derive actionable insights from data allows them to make informed decisions regarding revenue management, customer service, and operations.

**What is the impact of SQL, Python, and Power BI on enhancing operational efficiency within the hospitality industry?**

SQL and Python are instrumental in data preprocessing, analysis, and modelling tasks, enabling hospitality companies to extract valuable insights from datasets efficiently. Power BI, on the other hand, enhances operational efficiency by providing intuitive dashboards and reports that facilitate data-driven decision-making across various departments such as sales, marketing, and operations.

**Which important hospitality industry decision-making procedures stand to benefit most from data analytics?**

Decision-making procedures related to revenue management, customer segmentation, marketing strategies, and resource allocation stand to benefit significantly from data analytics in the hospitality industry. For example, analyzing booking patterns and cancellation policies can optimize revenue management strategies, while customer segmentation based on demographics and preferences can improve personalized marketing efforts.

**Which data analytics platform works best for particular analytical requirements in the hospitality sector?**

Power BI is a popular choice in the hospitality sector due to its user-friendly interface, extensive visualization capabilities, and integration with various data sources. However, the choice of platform depends on specific analytical requirements. For instance, SQL and Python are preferred for advanced data manipulation and modelling tasks, while Power BI excels in data visualization and dashboarding.

**In what ways can hospitality enterprises effectively incorporate data analytics into their more comprehensive strategic planning procedures?**

* Hospitality enterprises can incorporate data analytics into their strategic planning procedures by:
* Establishing clear objectives and key performance indicators (KPIs) aligned with business goals.
* Collecting and integrating data from various sources, including booking systems, customer feedback, and market trends.
* Analyzing historical data to identify patterns, trends, and potential areas for improvement.
* Implementing predictive analytics models to forecast demand, optimize pricing strategies, and mitigate risks.
* Continuously monitoring performance metrics and adjusting strategies based on real-time insights to stay agile and competitive in the market.