INFS3208 – Hotel Bookings Prediction Project
Proposal

Anmol Arora – 47945209

• Background

This project is about leveraging data analysis and machine learning techniques to gain insights and make predictions in the hotel industry. The goal is to enhance decision-making, optimize revenue, and improve customer service by understanding customer behavior and booking patterns.

• Motivation

This project is important because it empowers hotels to make informed decisions that lead to increased revenue, reduced costs, and improved customer satisfaction. It helps hotels stay competitive in a dynamic industry by leveraging the power of data analytics and machine learning.

• Objective and Features

The project has several key features, each serving a specific purpose:

1. Linear Regression Model (ADR Prediction):
- Objective: Predict the Average Daily Rate (ADR) of hotel bookings.
- Feature: Utilizes historical booking data, such as lead time and customer information, to build a predictive model for pricing.
2. Random Forest Model (Cancellation Prediction):
- Objective: Predict whether a booking will be canceled or not.
- Feature: Uses a machine learning algorithm to classify bookings as potential cancellations, helping hotels manage room availability efficiently.
3. Clustering Analysis (Customer Segmentation):
- Objective: Identify the most profitable client segments based on lead time and ADR.
- Feature: Applies clustering techniques to group customers with similar booking behaviors and characteristics, allowing hotels to target high-value segments.
4. FPGrowth Algorithm (Association Rule Mining):
- Objective: Discover association rules among different booking features.
- Feature: Applies the FPGrowth algorithm to uncover patterns and relationships between booking attributes, helping hotels tailor their services and amenities.

• Limitations of Traditional Computing

1. Processing Power: Traditional computers have finite processing power. They can only handle a limited number of calculations per second, which can be a significant bottleneck when dealing with large datasets or complex computations.
2. Memory: Traditional computers have limited memory (RAM). This restricts the amount of data that can be processed at once. Large datasets may not fit into memory,

leading to slower performance as data is constantly read from and written to disk.
3. Scalability: Traditional computing systems can be challenging to scale. Adding more processing power or memory to a single machine has its limits. To handle increasing workloads, organizations often need to invest in expensive hardware upgrades.
4. Speed: While traditional computers are fast, they may not be fast enough for real-time or near-real-time processing, especially in applications like high-frequency trading or real-time analytics.

- ## Benefits of Cloud Computing

1. Scalability: Cloud platforms allow you to scale your computing resources up or down based on project needs. This is particularly advantageous when dealing with varying workloads and large datasets in hotel prediction.
2. Cost-Efficiency: Cloud services often follow a pay-as-you-go model, which means you only pay for the computing resources you use. This can be more cost-effective than investing in and maintaining on-premises hardware.
3. Flexibility: Cloud platforms offer a wide range of services and tools that can be easily integrated into your project, allowing you to choose the best solutions for your specific needs.
4. Data Storage and Management: Cloud providers offer reliable and scalable data storage solutions, making it easier to store and manage the vast amounts of data typically associated with hotel prediction projects.

- ## Technologies Used & Cost Estimation

We harnessed cloud computing tools to make our predictions. Our data found a home in Hadoop Distributed File System (HDFS) and was consistently deployed using Docker Compose. Collaboration thrived on Jupyter Notebook, while PySpark for Apache Spark enabled distributed data processing and machine learning. These tools created a simplified and scalable environment for data exploration, model training, and predictive analytics.

The expected monthly cost for utilizing Google Cloud Platform (GCP) cloud resources in Sydney is AUD 118.58. This estimate includes a free operating system, a 30% sustained use discount, and a Compute Engine instance with a n1-standard-2 configuration running for 730 hours monthly. An additional 50 GiB auxiliary persistent disk contributes to the overall monthly cost. Please note that real costs may vary based on usage patterns and changes in GCP pricing. To ensure budget adherence when using GCP, prudent financial management is recommended.