Price Optimisation

1. Calculate Average Flight and Hotel Price by Destination

Mapper:

A computer screen shot of a program code

Description automatically generated

Reducer:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

Output:

A computer screen with white text

Description automatically generated

A computer screen with numbers and letters

Description automatically generated

The analysis highlights the variation in average flight and hotel prices across different destinations. Salvador (BH) emerges as the most expensive destination for both flight and hotel prices, with an average flight cost of 1179.23 RM and an average hotel price of 263.41 RM, likely reflecting its popularity or premium offerings. Florianopolis (SC) also shows high average flight prices at 1082.06 RM, with a moderately priced hotel stay averaging 313.02 RM. These figures suggest strong demand or premium positioning for these locations.

In contrast, Sao Paulo (SP) is the most affordable destination, with the lowest average flight price of 826.55 RM and an average hotel price of 139.10 RM, making it an attractive option for budget-conscious travelers. Similarly, Natal (RN) and Rio de Janeiro (RJ) offer relatively affordable flights, with prices averaging 866.97 RM and 893.07 RM, respectively, although Rio de Janeiro has slightly higher hotel costs at 165.99 RM compared to Natal’s 242.88 RM.

Destinations like Brasilia (DF) and Campo Grande (MS) fall into the mid-range category, with flight prices averaging 906.04 RM and 912.29 RM, respectively, and hotel prices at 247.62 RM and 260.39 RM. Recife (PE) balances affordability and demand, with an average flight price of 919.72 RM and hotel costs of 312.83 RM, making it a competitive option for travelers.

Overall, destinations with higher flight and hotel prices, such as Salvador and Florianopolis, indicate their appeal to premium travelers or high demand. Conversely, cities like Sao Paulo and Natal are positioned as budget-friendly options. Airlines and travel agencies can leverage these insights to create tailored promotions, such as bundled offers for premium destinations or discounts for budget-friendly ones, to maximize revenue and attract diverse traveler segments.

1. To analyze how flight prices vary by flightType

Mapper:

A computer screen with text

Description automatically generated

Reducer:

A computer screen shot of a program

Description automatically generated

A screenshot of a computer program

Description automatically generated

Output:

A computer screen with white text

Description automatically generated

The analysis reveals significant variations in flight prices based on the flight type. **Economic class** is the most affordable, with an average price of **660.90 RM**, catering primarily to budget-conscious travelers who prioritize cost savings over additional services. **Premium class** offers a balance between luxury and affordability, with an average price of **920.64 RM**, making it an appealing choice for business travelers or individuals seeking enhanced comfort without the high cost of first class. At the top of the spectrum, **first class** has the highest average price of **1181.55 RM**, reflecting its exclusivity and luxurious services, which appeal to affluent customers and those seeking a premium travel experience. These insights suggest that airlines can optimize their revenue by targeting different customer segments: offering upsells for economic travelers, promoting premium class as a value-for-money option, and focusing on exclusivity and high-end services for first-class passengers.

1. Price Optimization Based on Flight Demand

Analyze flight prices by from-to routes to identify high-demand routes.

A computer screen with text and images

Description automatically generated

A computer screen shot of a program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The output provides insights into flight demand and pricing for various routes. Routes are represented as from-to combinations, with the data showing the average flight price and total number of flights for each route. High-demand routes, such as Aracaju (SE)-Florianopolis (SC) (2610 flights) and Florianopolis (SC)-Aracaju (SE) (2610 flights), tend to have higher average prices, like 1227.86 RM and 1240.04 RM, respectively. In contrast, shorter routes or routes with lower demand, such as Natal (RN)-Rio de Janeiro (RJ) (283 flights), have higher price variability, with an average price of 1147.01 RM.

Overall, shorter routes (e.g., Sao Paulo (SP)-Rio de Janeiro (RJ) with 276 flights at 875.36 RM) are generally priced lower, while longer routes (e.g., Florianopolis (SC)-Natal (RN) with 2029 flights at 1255.50 RM) have higher prices. This pattern reflects typical airline pricing strategies, where distance and demand influence ticket costs. High-demand routes present opportunities for price optimization, while low-demand routes may benefit from targeted promotions to boost bookings.

1. Seasonal Price Trends

Objective: Identify peak travel months and average prices during those months:✅

A computer screen shot of a program code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A computer screen shot of a computer program

Description automatically generated

A computer screen with white text

Description automatically generated

The data provides insights into the peak travel months based on the number of flights and their corresponding average flight prices. December (month 12) emerges as the peak travel month, with 7873 flights and an average flight price of 957.36 RM, followed closely by October (month 10) with 8663 flights and an average price of 962.56 RM. November (month 11) also shows high demand, with 7669 flights and an average price of 956.63 RM. These months align with holiday seasons and increased travel activity.

Conversely, months like February (month 2) and August (month 8) observe lower travel volumes, with 6659 and 5420 flights, respectively, and average prices ranging between 953.99 RM and 957.37 RM. Interestingly, prices remain relatively stable across months, averaging between 945.59 RM (January) and 966.21 RM (June), indicating consistent pricing strategies regardless of demand fluctuations.

This analysis highlights opportunities for airlines to maximize revenue by introducing targeted promotions during low-demand months while maintaining premium pricing during peak seasons. Additionally, price differentiation strategies for months with stable pricing could further optimize profitability.

Marketing Strategies:

October and December are high-demand months. Airlines and travel agencies should focus on maximizing revenue during these months with competitive pricing and promotions. August and June have lower demand, but June has higher prices. Consider offering discounts or promotional packages to attract more travelers in these months. January shows a balance of high demand and lower prices. This indicates an opportunity to increase revenue through modest price adjustments without losing travelers.

Focus on family and holiday travel promotions in December and October.

Offer discounts and packages for summer travel (June-August) to boost demand during off-peak periods.

Distance-Based Price Analysis

1. To analyze how **flight prices vary by flight type and distance**,

Mapper:

A computer screen shot of a program code

Description automatically generated

Reducer

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

The analysis reveals how flight prices vary by flight type (economic, premium, and first-class) and travel distance (short, medium, and long). Economic flights are the most affordable across all distance categories, with short-distance flights (≤300 km) averaging 427.16 RM, medium-distance flights (301–800 km) averaging 676.79 RM, and long-distance flights (>800 km) averaging 886.30 RM. Medium-distance economic flights are the most popular, with 14800 flights, reflecting strong demand for affordable options over moderate distances.

First-class flights are the most expensive, particularly for long-distance travel, averaging 1595.31 RM with a maximum of 1754.17 RM across 5504 flights. Medium-distance first-class flights are also popular, with 22250 flights averaging 1210.97 RM, making them the most frequently chosen first-class option. However, first-class short-distance flights (≤300 km) are less popular, averaging 753.77 RM across 6854 flights, likely due to limited perceived value for shorter trips.

Premium flights balance cost and comfort, making them a popular choice for medium-distance routes, with 14884 flights averaging 942.25 RM. Long-distance premium flights average 1253.30 RM, appealing to travelers seeking luxury at a lower price point than first class, while short-distance premium flights average 580.64 RM, making them the least frequent among premium options with just 4646 flights.

Overall, economic flights dominate medium-distance travel due to their affordability, while first-class long-distance flights cater to premium customers. Airlines can optimize pricing by offering discounts on underutilized premium short-distance flights and promoting economic options for medium-distance travel to maintain market share. Additionally, pricing strategies for first-class long-distance flights can be adjusted to capitalize on their steady demand.

Time-of-Day Pricing

**Objective: Analyze flight prices by time.**

A computer screen with text

Description automatically generated

A computer screen shot of a program code

Description automatically generated

A screenshot of a computer program

Description automatically generated

Airlines can enhance profitability by targeting underutilized premium short-distance flights with discounts and promoting economy fares for medium distances. Adjusting pricing strategies for long-distance first-class flights can also maximize returns on consistent demand.A screenshot of a computer

Description automatically generated

The data provides insights into how flight prices vary based on the time of day (in decimal hours). Flights departing during **early hours of the day** (e.g., **0.44 - 0.86** hours, corresponding to approximately 12:26 AM to 12:51 AM) have the lowest average prices, ranging from **486.88 RM** to **828.67 RM**, likely due to reduced demand during late-night and early-morning hours. As the day progresses, prices generally increase, peaking during **1.66 - 2.44 hours** (1:39 AM to 2:26 AM), with prices reaching as high as **1347.46 RM**.

The most expensive flights are clustered around **early morning hours**, such as **1.76 hours** (1:46 AM, **1077.35 RM**) and **2.44 hours** (2:26 AM, **1347.46 RM**), reflecting higher demand for morning departures. These flights are often preferred by business travelers and early commuters. In contrast, flights departing later in the evening or early at night tend to be priced more affordably. This trend highlights an opportunity for airlines to optimize pricing by offering discounts during late-night or early-morning slots to fill seats while maintaining premium pricing for high-demand morning hours

7. Price Variability by Booking Agency (New).

A computer screen shot of a program code

Description automatically generated

A computer screen shot of a program code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A computer screen shot of a cat

Description automatically generated

The analysis of price variability by booking agency reveals significant differences in average flight and hotel prices among agencies. **FlyingDrops** offers the highest average flight price at **1186.13 RM** and an average hotel price of **539.72 RM**, positioning itself as a premium agency catering to high-end travelers seeking luxury services. In contrast, **CloudFy** provides more affordable options, with an average flight price of **917.02 RM** and a hotel price of **534.99 RM**, appealing to budget-conscious travelers. **Rainbow** falls in the mid-range category, offering an average flight price of **922.96 RM** and a hotel price of **536.32 RM**, striking a balance between affordability and quality.

These insights highlight the importance of pricing strategies in attracting different customer segments. Airlines and hotels partnering with these agencies can tailor their offerings to align with the agency's target market. For instance, FlyingDrops could focus on exclusive packages for premium travelers, while CloudFy could leverage its affordability to attract price-sensitive customers. By understanding these pricing dynamics, agencies can better position themselves in a competitive travel market.

Comparison performance mapreduce and Hive.   
1.

|  |  |  |
| --- | --- | --- |
|  | Hive | MapReduce |
| 1. | Query ID = manis\_20250102131705\_4c0bfc77-0dc1-43b6-bdd9-35214f7aa61e  Total jobs = 2  Launching Job 1 out of 2  Number of reduce tasks not specified. Estimated from input data size: 1  In order to change the average load for a reducer (in bytes):  set hive.exec.reducers.bytes.per.reducer=<number>  In order to limit the maximum number of reducers:  set hive.exec.reducers.max=<number>  In order to set a constant number of reducers:  set mapreduce.job.reduces=<number>  Starting Job = job\_1735710072368\_0023, Tracking URL = http://manis-VirtualBox:8088/proxy/application\_1735710072368\_0023/  Kill Command = /home/manis/hadoop/bin/mapred job -kill job\_1735710072368\_0023  Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  2025-01-02 13:17:19,656 Stage-1 map = 0%, reduce = 0%  2025-01-02 13:17:28,205 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.53 sec  2025-01-02 13:17:37,772 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.13 sec  MapReduce Total cumulative CPU time: 9 seconds 130 msec  Ended Job = job\_1735710072368\_0023  Launching Job 2 out of 2  Number of reduce tasks determined at compile time: 1  In order to change the average load for a reducer (in bytes):  set hive.exec.reducers.bytes.per.reducer=<number>  In order to limit the maximum number of reducers:  set hive.exec.reducers.max=<number>  In order to set a constant number of reducers:  set mapreduce.job.reduces=<number>  Starting Job = job\_1735710072368\_0024, Tracking URL = http://manis-VirtualBox:8088/proxy/application\_1735710072368\_0024/  Kill Command = /home/manis/hadoop/bin/mapred job -kill job\_1735710072368\_0024  Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1  2025-01-02 13:17:54,767 Stage-2 map = 0%, reduce = 0%  2025-01-02 13:18:02,177 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec  2025-01-02 13:18:10,552 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.77 sec  MapReduce Total cumulative CPU time: 5 seconds 770 msec  Ended Job = job\_1735710072368\_0024  MapReduce Jobs Launched:  Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.13 sec HDFS Read: 13467419 HDFS Write: 529 SUCCESS  Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.77 sec HDFS Read: 8296 HDFS Write: 458 SUCCESS  Total MapReduce CPU Time Spent: 14 seconds 900 msec  OK | packageJobJar: [/home/manis/mapper.py, /home/manis/reducer.py, /tmp/hadoop-unjar3846762099075013372/] [] /tmp/streamjob5888350201400115758.jar tmpDir=null  2025-01-02 13:00:46,648 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032  2025-01-02 13:00:46,993 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032  2025-01-02 13:00:47,351 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/manis/.staging/job\_1735710072368\_0022  2025-01-02 13:00:47,983 INFO mapred.FileInputFormat: Total input files to process : 1  2025-01-02 13:00:48,089 INFO mapreduce.JobSubmitter: number of splits:2  2025-01-02 13:00:48,304 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1735710072368\_0022  2025-01-02 13:00:48,305 INFO mapreduce.JobSubmitter: Executing with tokens: []  2025-01-02 13:00:48,641 INFO conf.Configuration: resource-types.xml not found  2025-01-02 13:00:48,641 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.  2025-01-02 13:00:48,766 INFO impl.YarnClientImpl: Submitted application application\_1735710072368\_0022  2025-01-02 13:00:48,827 INFO mapreduce.Job: The url to track the job: http://manis-VirtualBox:8088/proxy/application\_1735710072368\_0022/  2025-01-02 13:00:48,830 INFO mapreduce.Job: Running job: job\_1735710072368\_0022  2025-01-02 13:00:58,085 INFO mapreduce.Job: Job job\_1735710072368\_0022 running in uber mode : false  2025-01-02 13:00:58,087 INFO mapreduce.Job: map 0% reduce 0%  2025-01-02 13:01:08,530 INFO mapreduce.Job: map 50% reduce 0%  2025-01-02 13:01:09,543 INFO mapreduce.Job: map 100% reduce 0%  2025-01-02 13:01:16,659 INFO mapreduce.Job: map 100% reduce 100%  2025-01-02 13:01:17,679 INFO mapreduce.Job: Job job\_1735710072368\_0022 completed successfully  2025-01-02 13:01:17,882 INFO mapreduce.Job: Counters: 55  File System Counters  FILE: Number of bytes read=2521428  FILE: Number of bytes written=5980304  FILE: Number of read operations=0  FILE: Number of large read operations=0  FILE: Number of write operations=0  HDFS: Number of bytes read=13451526  HDFS: Number of bytes written=263  HDFS: Number of read operations=11  HDFS: Number of large read operations=0  HDFS: Number of write operations=2  HDFS: Number of bytes read erasure-coded=0  Job Counters  Killed map tasks=1  Launched map tasks=2  Launched reduce tasks=1  Data-local map tasks=2  Total time spent by all maps in occupied slots (ms)=32384  Total time spent by all reduces in occupied slots (ms)=10876  Total time spent by all map tasks (ms)=16192  Total time spent by all reduce tasks (ms)=5438  Total vcore-milliseconds taken by all map tasks=16192  Total vcore-milliseconds taken by all reduce tasks=5438  Total megabyte-milliseconds taken by all map tasks=4145152  Total megabyte-milliseconds taken by all reduce tasks=1392128  Map-Reduce Framework  Map input records=81105  Map output records=81104  Map output bytes=2359214  Map output materialized bytes=2521434  Input split bytes=224  Combine input records=0  Combine output records=0  Reduce input groups=9  Reduce shuffle bytes=2521434  Reduce input records=81104  Reduce output records=9  Spilled Records=162208  Shuffled Maps =2  Failed Shuffles=0  Merged Map outputs=2  GC time elapsed (ms)=421  CPU time spent (ms)=7060  Physical memory (bytes) snapshot=753659904  Virtual memory (bytes) snapshot=5608611840  Total committed heap usage (bytes)=512229376  Peak Map Physical memory (bytes)=291209216  Peak Map Virtual memory (bytes)=1866936320  Peak Reduce Physical memory (bytes)=182448128  Peak Reduce Virtual memory (bytes)=1874874368  Shuffle Errors  BAD\_ID=0  CONNECTION=0  IO\_ERROR=0  WRONG\_LENGTH=0  WRONG\_MAP=0  WRONG\_REDUCE=0  File Input Format Counters  Bytes Read=13451302  File Output Format Counters  Bytes Written=263  2025-01-02 13:01:17,882 INFO streaming.StreamJob: Output directory: /manis/hadoop/output\_hive\_query |
|  |  |  |