Passport Paradise: Your Gateway to the World

Isabelle Stuadt

Haylee Hidlebaugh

Abigail Cobb

Alizabeth Recker

Kairavi Dave

1127 Gerdin 2167 Union Drive Ames Iowa

p. 515-294-5556

passportparadise.com

passportparadise@gmail.com

Table of Contents

Executive Summary

Passport paradise is a travel company specializing in crafting personalized and unforgettable journeys for clients. By analyzing extensive travel data, we deliver insights that enable us to curate bespoke itineraries aligned with individual preferences and trends. Leveraging cutting-edge data analytics, our approach ensures eco-friendly travel options, seamless planning, and exceptional customer satisfaction.

This project provides the following data-driven insights:

➢ Analysis of user preferences to identify trends in destinations, travel seasons, and customer behaviors

➢ Integration of data from flights, hotels, and user reviews to create a comprehensive view of travel patterns

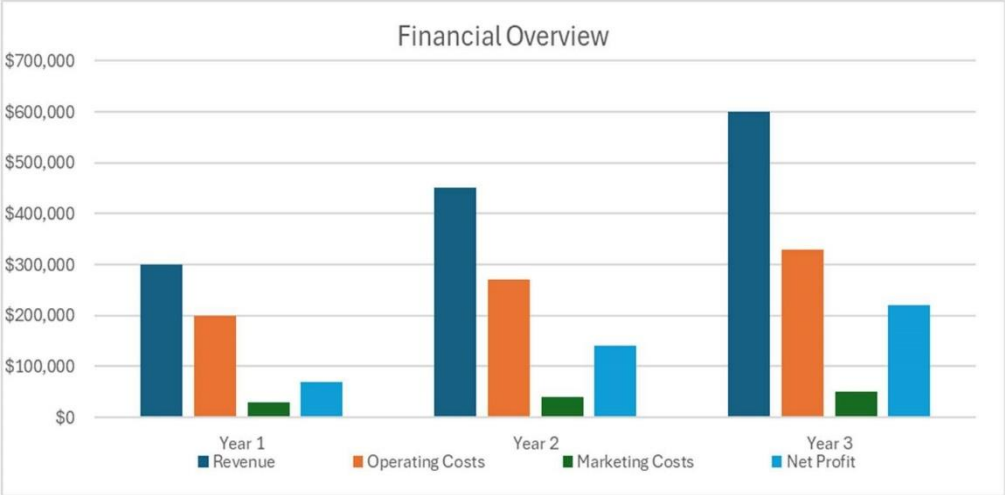
➢ Predictive modeling to offer tailored recommendations and exclusive travel deals based on customer interests

➢ Visualizations and dashboards showcasing customer segments, booking trends, and seasonality for strategic planning

Highlights

The target market for the Passport Paradise is strategically segmented to cater to a diverse range of travelers, ensuring broad market coverage and revenue potential. By catering to a diverse range of customer segments, from adventure seekers to eco-tourist travelers, Passport paradise positions itself to serve the broad and dynamic travel industry. Each segment presents unique opportunities for growth, from high-value luxury tourists to niche markets like eco-tourism and cultural travelers. This diverse customer base not only spreads risk but also allows the business to capitalize on emerging trends within the global

travel market. By offering tailored packages that meet the specific needs of each segment, passport paradise can build a strong reputation and ensure long-term customer loyalty.



The projected financial overview chart is crucial for Passport Paradise, providing critical insights into the company's expected financial trajectory over the next three years. It allows Passport Paradise to anticipate revenue growth, monitor operating and marketing costs, and forecast net profits, which are critical for strategic planning. By understanding these projections, Passport Paradise can make informed decisions about resource allocation, identify areas where costs can be further optimized, and plan for future investments in marketing and operations.

Objectives

- Identify meaningful trends and patterns from data gathered on travel experiences, reviews, and destination preferences
- Develop strategies focused on marketing, identifying target audience, and optimizing travel experience
- Use data analysis to describe data, make predictions, and develop suggestions to the travel industry

Keys to Success

Data-Driven Travel Insights: We leverage data analytics to identify trends in traveler preferences, experiences, and reviews, providing actionable insights to optimize marketing strategies and enhance the customer journey.

Targeted Marketing Strategies: By analyzing customer data, we identify key audience segments, allowing for personalized marketing campaigns that resonate with travelers' specific interests and destination preferences.

Predictive Analytics for Travel: Utilizing advanced data techniques, we forecast emerging travel trends and preferences, enabling businesses to stay ahead of the curve and tailor offerings to future demands.

Enhanced Travel Experience: Our continuous focus on improving customer satisfaction ensures that we use data to suggest and implement meaningful improvements in the travel experience.

Description of Business

Our Travel company is committed to creating individualized travel experiences that meet each client's requirements and interests. Our unique itineraries, exclusive travel offers, and round-the-clock customer service set us apart and guarantee a smooth and stress-free travel experience for our customers. Our distinctive method consists of customized travel itineraries created by knowledgeable advisors, exclusive savings obtained through solid industry relationships, and 24hour support for any concerns pertaining to travel. Our main objectives are to encourage ecofriendly travel and attain exceptional customer happiness. By concentrating on these locations, we hope to emerge as the leading option for tourists looking for unique and unforgettable experiences.

Products and Services

- Customers are matched with a personal experience curator who specializes in designing itineraries that are specifically catered to their interest.
- Create a custom app for each trip that includes interactive maps, daily schedules, realtime updates, and local recommendations. The app could also have built-in features like language translation and emergency contacts.
- Create a subscription service that provides members with regular travel inspiration, exclusive offers, early access to deals, and curated travel content tailored to their interests.
- Provide clients with special rates, upgrades, or added amenities at luxury hotels and resorts.
- Offer integrated booking services that include flights, accommodations, and ground transportation.

Our main objective is to give customers amazing travel experiences by providing professional guidance, customized itineraries, and easy booking services. This involves figuring out the practicalities of the trip, getting to know the client's preferences, and making sure the trip is easy and pleasurable.

Financial Data About the Company

In the world of travel, smart financial management is key to ensuring our agency’s success. In this section, we will explore the financial strategies that will help us maintain profitability and build a sustainable business. We’ll discuss how we plan to manage our cash flow effectively,

optimize our operating costs, and outline the financial requirements necessary to keep our travel agency running smoothly.

The core of ensuring our travel agency's profitability lies in effective cash flow management, which includes the following components:

Strategic Cash Flow Management: Given the initial investment of $20,000 in equipment and software, along with legal and setup costs, we will ensure steady cash flow by securing upfront deposits for travel bookings and tours. By requiring clients to pay a portion in advance, we reduce the risk of late or missed payments and maintain liquidity for ongoing operations like rent, utilities, and payroll.

Targeted Marketing and Advertising ROI: With $8,000 allocated for advertising and promotions, we will focus on cost-effective digital marketing strategies, such as social media advertising and search engine optimization (SEO), to attract travelers who are searching for personalized travel experiences. Tracking the ROI from these efforts ensures we maximize every marketing dollar, reaching the right audience without overspending.

Client-Centered Payment Terms: To reduce the impact of late payments, we will create clear payment terms that require an initial deposit upon booking and full payment before travel. This minimizes the risk of cash flow issues and ensures we can meet operating costs like payroll and utilities, which total $50,000 and $2,500 per month respectively.

Optimizing Operating Costs: While our rent in Des Moines, Iowa is projected at $18,000 per year, we will aim to minimize utility expenses and other operational costs by implementing energy-efficient practices and negotiating better service agreements for utilities and office supplies. Keeping overhead costs low will allow us to invest in areas that directly contribute to client satisfaction.

Leveraging Partnerships: We will work closely with hotels, airlines, and local travel partners to negotiate discounted rates, increasing our margins while offering competitive pricing to customers. These partnerships can also help us build exclusive travel experiences that attract repeat customers.

Start Up Costs and Operating Budget:

- *Equipment and Supplies*: We estimate $20,000 to be allocated towards initial equipment and software
- *Legal and Professional:* Legal fees and business establishment cost us $5,000
- *Renovation/Design:* As we will be meeting with clients in the office, it is important that we create a welcoming environment. We have reserved $10,000 for renovations and office set up. Advertising and Promotion: We have allocated $8,000 towards start-up promotion and website development.

- *Rent:* We will have 5 full-time employees and will be located in Des Moines, Iowa. A standard office space for this many people in Des Moines costs about $18,000 per year.
- *Utilities:* We estimate for an office space of about 1,500 square feet to spend $2,500 per month on utilities such as heating and cooling, water, electricity, and internet.
- *Insurance:* We expect to spend $500 per month on health insurance, and $100 per month plus a $10,000 deductible for professional indemnity insurance.
- *Payroll:* Salary and incentives will cost $50,000 per month.
- *Loan Payments:* We will take out a loan of $40,000 which will require monthly payments of $500 to be paid off in ten years
- *Office Supplies:* We have devoted $3,000 for office supplies per month
- *Travel and Entertainment:* $500 per month will be reserved for company outings, entertainment, and work-related travel.
- *Legal and Accounting:* We estimate that compliance and accounting costs will total $6,000 per year.
- *Advertising and Promotion:* We will allocate $8,000 per year for advertising
- *Repairs and Maintenance:* We budget $2,000 per month for miscellaneous repairs and maintenance.
- *Depreciation:* $4,000 will designated for depreciation

Project Detail

Analytical Questions

The dataset we are using for this project is from Kaggle, an online platform that supports data science and machine learning efforts by offering datasets, competitions, and tutorials for data exploration and modeling.

Specifically, we are using the Argo Datathon 2019 dataset, which contains synthetic data related to corporate travel, including trip times, distances, durations, and transportation modes. This dataset simulates real-world travel data, making it valuable for testing machine learning models that optimize travel behavior and predictions within the business travel sector. Key subdivisions in this dataset include:

- Trip Details: Information on departure and arrival locations, times, and distances.
- Transportation Modes: Types of transport used for trips, such as air travel, car rentals, and public transport.
- Travel Durations: Metrics that detail how long each trip takes.
- Cost Data: Information regarding the expenses incurred during travel.

This big data project aims to create a model using semi-structured data in order to establish recommendations for changes and additions from a business standpoint.

Type of Analytics used:

- Descriptive Analytics: This looks at booking patterns, customer types, and travel choices. By making reports and charts, we can see what customers like and popular places, helping us improve marketing and operations.
- Diagnostics Analytics: By finding patterns, we can understand what affects customer satisfaction, the success of marketing, and changes in travel preferences. This helps us make better decisions and improve services for a better customer experience.
- Predictive Analytics: This predicts travel trends, peak demand times, and possible issues. It helps us make smart choices about marketing, pricing, and resources, keeping us competitive in a changing market.
- Prescriptive Analytics: This involves adjusting prices for flights and hotels based on current demand and customer behavior. It also includes creating personalized marketing campaigns to attract specific customer groups.

Research Questions

In this section, we outline the key research questions that guide our exploration of consumer behavior and preferences within the travel industry, leveraging big data analytics to enhance customer experiences and optimize business strategies.

Demographics:

1. What Demographics factors correlate with specific outcomes?
    a. This will help us understand how certain demographics influence can guide choices that travelers are more likely to make
2. What factors influence travelers' destination choices
    a. Allows us to evaluate key influences such as cost, culture, climate with allow us to make sure we are using the correct marketing strategy tailor to our business

Trend:

3. Are there significant trends over time in the data?
    a. Identifying trends can help forecast future outcomes and evaluate the effectiveness of specific programs and policies.
4. What are the seasonal patterns in destination popularity?
    a. Analyzing seasonal trends can help in resource planning, pricing strategies, and promotional campaigns

Predictive:

5. What are the most predictive variables for the outcomes?
    a. Focusing on specific and predictive variables will enhance decision making and allocation

Data Sources and Acquisition

Data Source: We sourced our datasets from Kaggle, which includes files on Flights, Hotels, and Users for our analysis. https://www.kaggle.com/datasets/leomauro/argodatathon2019

- You must have a Kaggle account to access and download the dataset. If you don't have one, sign up or log in.

- Once logged in, navigate to the "Data" tab on the dataset page.

The dataset is composed of multiple JSON files, each representing different types of data such as business information, user reviews, user profiles, check-ins, tips, and photos. A summary of these JSON files and their respective number of attributes is provided below:

Flights Dataset (flights.json)

Description: Contains information about flight bookings, including details on routes, prices, and agencies. Number of Attributes: 8 Key Attributes:

- travelCode: integer (identifier for travel instance)
- userCode: integer (identifier for the user booking the flight)
- from: string (departure city)
- to: string (arrival city)
- flightType: string (class of service, e.g., firstClass)
- price: float (cost of the flight)
- time: float (duration of the flight in hours)
- distance: float (distance of the flight in kilometers)
- agency: string (travel agency providing the flight)
- date: string (date of the flight, formatted as MM/DD/YYYY)

Hotels Dataset (hotels.json)

Description: Contains information about hotel bookings, including details on prices, duration of stay, and locations. Number of Attributes: 7 Key Attributes:

- travelCode: integer (identifier for travel instance)
- userCode: integer (identifier for the user booking the hotel)
- name: string (name of the hotel)
- place: string (location of the hotel)
- days: integer (number of nights stayed)
- price: float (cost per night)
- total: float (total cost for the stay)
- date: string (date of the booking, formatted as MM/DD/YYYY)

Users Dataset (users.json)

Description: Contains detailed profiles of users, highlighting their characteristics and booking behaviors. Number of Attributes: 5 Key Attributes:

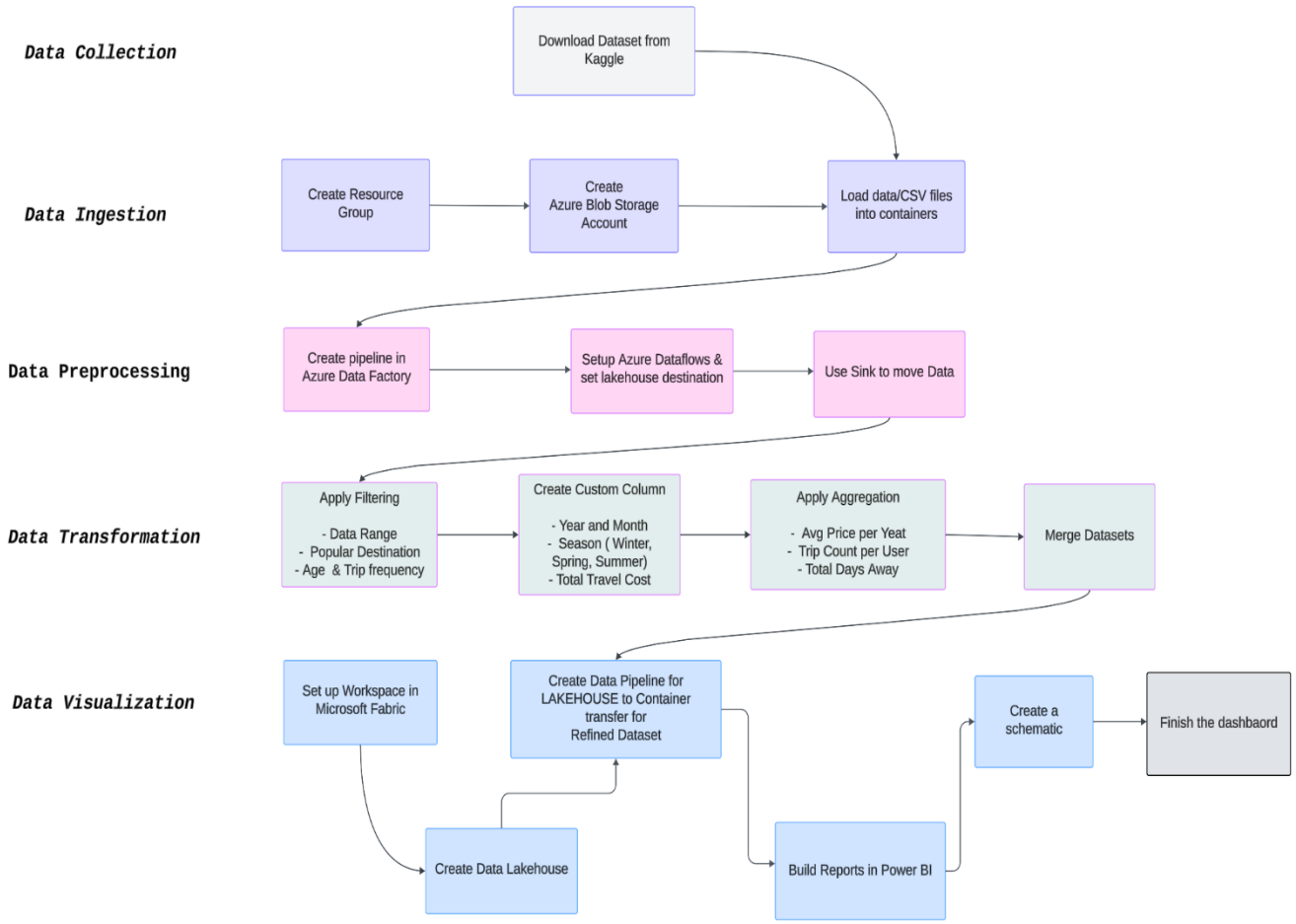
- user_id: integer (primary key, unique identifier for each user)
- company name: string (name of the user's company)
- name: string (name of the user)
- gender: string (gender of the user) age: integer (age of the user)

Tools for data Preparation

1. Data Cleaning and Preparation
    a. Power BI: Can use this to find specific trends and most popular decisions that travelers make when looking to book a trip. Which would include hotels, flights, and destinations.
    b. Excel: We can use this to break down out data sets into certain categories to find what each traveler/hotel/designation have in common
    c. Microsoft Fabric: This can easily connect to different data sources to gather and transform data in one place. Also uses power query to clean the data which. May include adding missing values, formatting issues, and merging datasets

2. Data Visualization
    a. Power BI: Useful for business intelligence and visual analytics to create graphs and queries to make the most popular trends knows
    b. Microsoft Fabric: Create dashboards that visualize key metrics such as, average hotel prices by season, popular destinations, or climate impacts on travel

Data Preparation

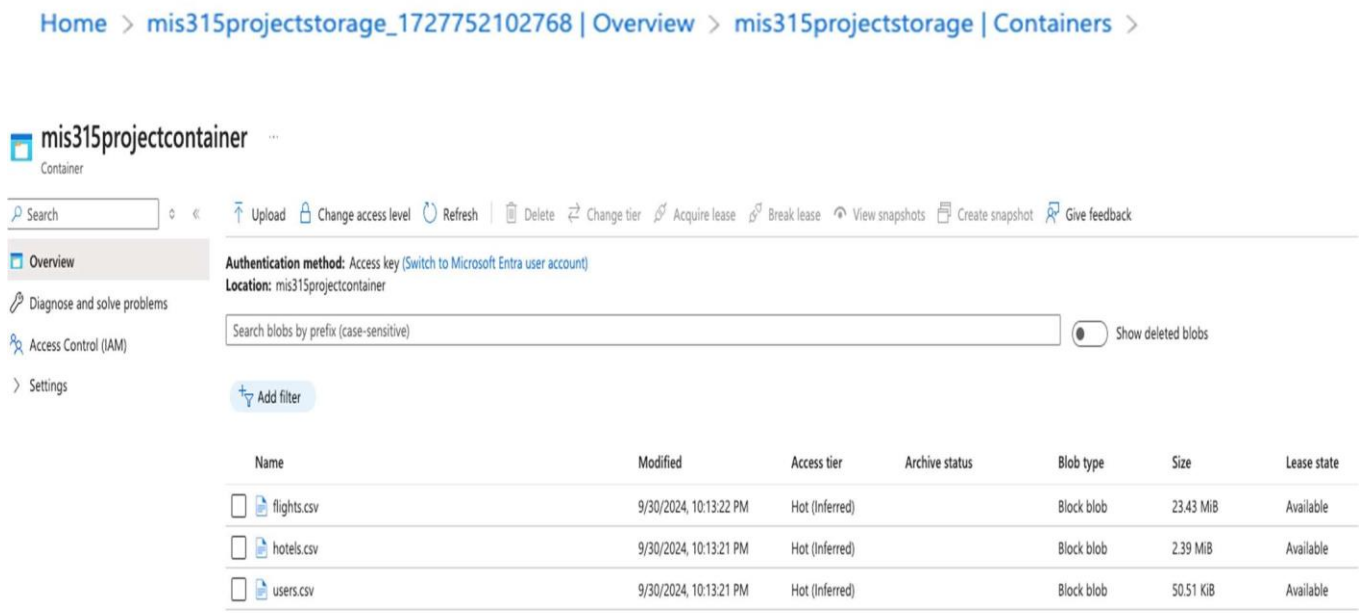


Data Ingestion & Storage in Azure

Create Resource Group: We started by creating a Resource Group in Azure to organize and manage all related resources for our project. This allows for better resource management and cost tracking as all components are grouped logically.

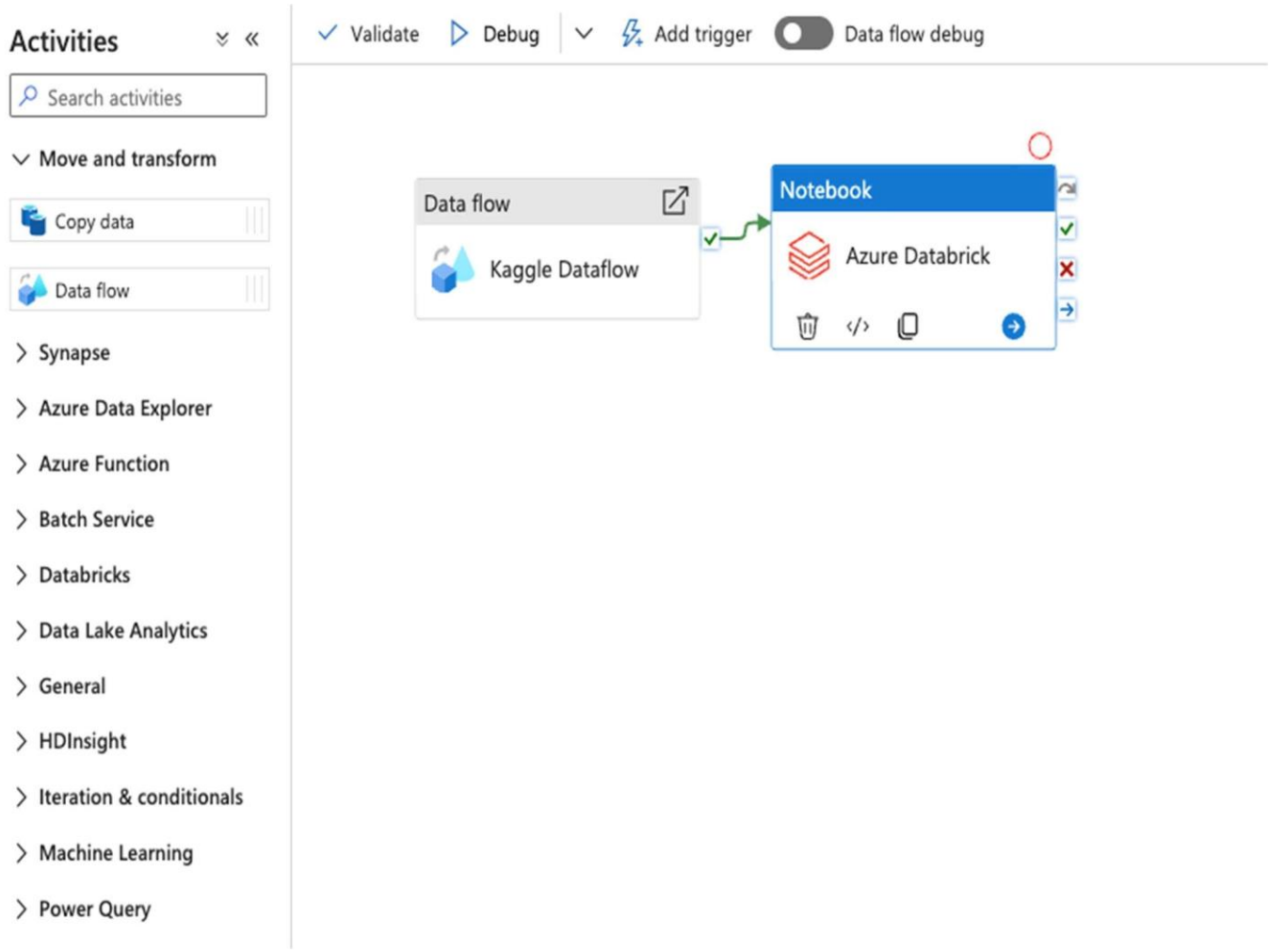
Create Storage Account: Following the resource group, we set up a Storage Account. This account will store our data in various containers, facilitating easy access and management of our raw and processed data.

Load Data into Containers: We then loaded the raw data into designated container within the storage account. These containers act as staging areas for our data, where it can be easily accessed for transformation and processing.



Data Processing in Azure

After loading the data, we will create a pipeline in Azure Data Factory (ADF). This pipeline will include a Dataflow for data transformation, where we can perform essential operations like data type conversions, filtering, and basic aggregations. To enhance our processing capabilities, we will integrate Azure Databricks within the ADF pipeline, allowing us to leverage Spark for advanced data manipulation and machine learning tasks.



The initial phase of our analysis involved creating a dataflow that facilitated column selection and data filtering to ensure we retained only the relevant information. We utilized three CSV files as our data sources: Flights, Hotels, and Users. Each dataset underwent individual selection and filtering processes tailored to extract essential columns and criteria. Afterward, we joined the results from these datasets using a join function to create a comprehensive view of our data, which would inform our visualizations and analyses in Power BI.

Overview of the Select and Filter usage on the datasets:

Data File	Select Columns	Filter
Flights	TravelCode, userCode, From, To, Flight type, Price, Time, Distance, Agency, Date	-Only include flights booked between 2019 and 2023 -Select flights with a price below the average price for that year.
Hotels	TravelCode, userCode, Name, Place, Days, Price, Total, Date	-Only include hotel stays from 2019 to 2023. - Filter hotels based on location to include only those in popular travel destinations. - Include stays where total price is below a specific threshold.
Users	Code, Company, Name, Gender, Age	- Only users with more than one trip booked are included. - Filter by age to focus on users aged 18-65. - Include users from specific companies or industries that show higher engagement.

We applied filters to clean and prepare the datasets by removing unnecessary columns and irrelevant records.

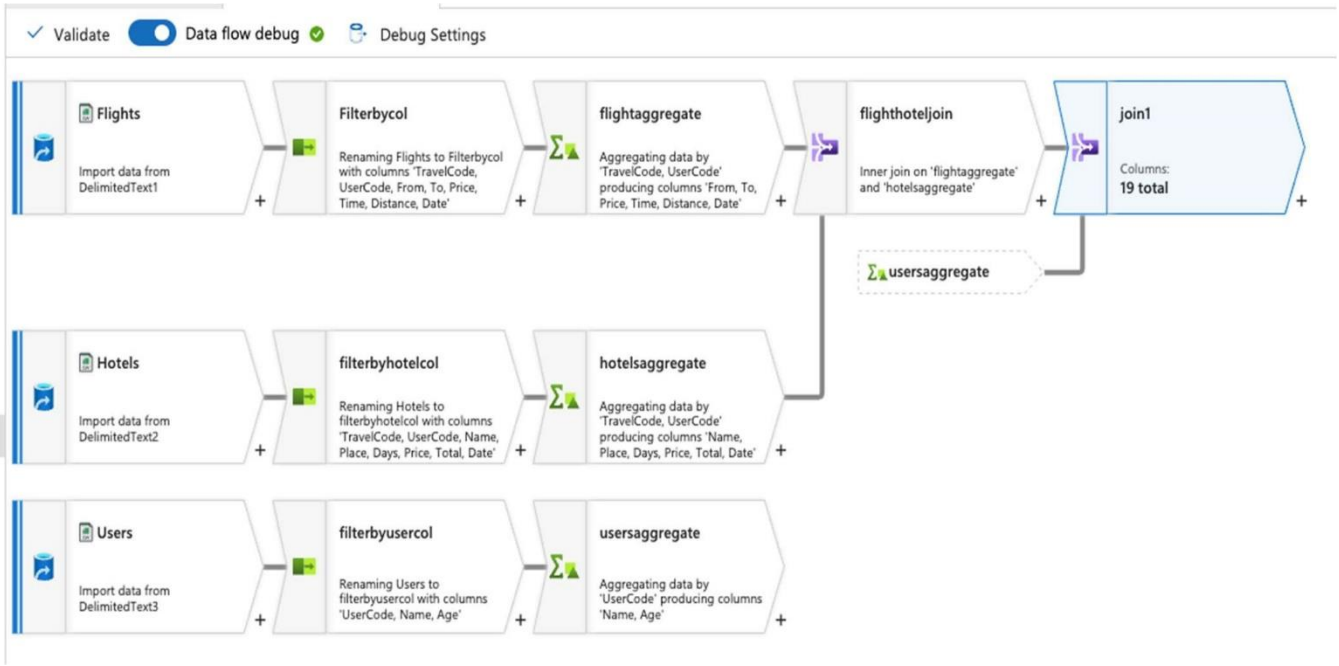
In the Flights dataset, we eliminated the Agency and Flight Type columns, as they were redundant and did not contribute valuable insights for analyzing travel patterns or user behavior. We also filtered out incomplete records, focusing on crucial fields like TravelCode, UserCode, and Price to ensure data integrity.

Similarly, in the Users dataset, we removed the Company and Gender columns, as they did not provide significant insights for our travel analysis.

After filtering, we identified and removed duplicate entries across the datasets to ensure a clean dataset for analysis. For the Flights dataset, duplicates were removed based on the TravelCode and UserCode columns, which serve as unique identifiers for each record. In the Hotels dataset, we applied the same criteria, ensuring each record was distinct. For the Users dataset, duplicates were eliminated using the unique identifier in the UserCode column.

The next step involved joining the datasets on common keys to the information into a single dataset for analysis. We joined the Flights and Hotels datasets using TravelCode and UserCode,

aligning travel bookings with hotel stays to create a comprehensive travel itinerary. Then, we joined the Users dataset with the combined Flights-Hotels dataset using UserCode, enriching the data with user-specific information such as name and age. We executed these joins using an inner join to ensure that only valid and complete records were carried forward for further analysis.



Data Lakehouse in Microsoft Fabric

To integrate with Microsoft Fabric, we will set up a workspace and create a Data Lakehouse. This will involve establishing an ADF pipeline that transfers data from our Azure SQL Database to the Lakehouse, ensuring a seamless flow of information.

Assumptions Made

1. Data collections and Bias in collection Method:

- Simple Random Sample:
    - Ideally, we presume that each dataset is a simple random sample of consumers, hotels, or flights. This indicates that there is an equal possibility for every item to be part of the dataset. The analysis might not be representative of the whole population if the sample procedure was skewed (for example, by concentrating on well-traveled routes, upscale lodgings, or frequent users).
- Non- Response/selection Bias:

- The analysis might be skewed by systematic exclusions, such as users who make few reservations or hotels in remote locations. We must think about whether some entry categories (users, hotels, or flights) may be over- or underrepresented.

2. Variance:

- Homogeneity:
  - Many statistical tests presume that the variances of several groups (such as different airlines, hotels, or user demographics) are similar. The credibility of the analysis may be impacted if variations are very different, particularly across significant categories (such as hotels in various locations or flights on various days).
- Independence:
  - We usually presume that each element in datasets like this is independent of the others. For instance, the booking habits of one user shouldn't affect those of another, nor should the delay of one flight have a direct impact on another's. Analysis may be impacted if this assumption is broken, for example, by associated delays between connected planes.

3. Sample Size

- Sufficiency:
  - o Ideally, each dataset (flights, hotels, users) should be substantial enough to capture the heterogeneity within the population it represents. Results from small samples may not be trustworthy, particularly for underrepresented populations.
- Adequate Representation o If the analysis involves comparing subgroups (e.g., hotels in different cities, users by age group), each subgroup should have a sufficient number of observations to ensure reliable conclusions.

4. Normality

- Many statistical tests assume normality (especially when analyzing averages or conducting hypothesis testing). For example, we might assume that flight delay times or hotel ratings follow a normal distribution. If the data is heavily skewed, it may require transformation or the use of non-parametric tests.

5. Stationarity (for Time-series data)

- If analyzing trends over time (e.g., flight delays by month or hotel occupancy over seasons), we assume the patterns are stable unless accounting for seasonality. Nonstationary data can lead to misleading interpretations if time-based trends are ignored.

Data Visualizations

Power BI: Useful for business intelligence and visual analytics to create graphs and queries to make the most popular trends knows

Microsoft Fabric: Create dashboards that visualize key metrics such as, average hotel prices by season, popular destinations, or climate impacts on travel

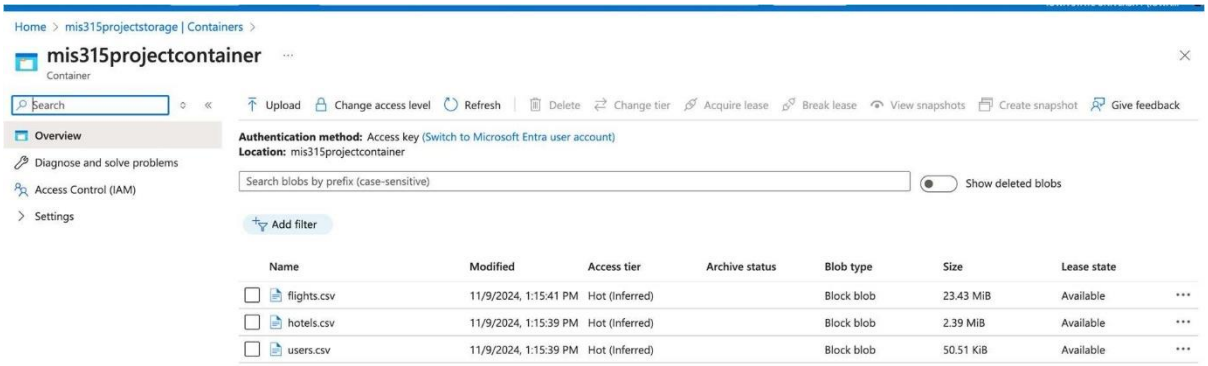
Data Process

We implemented a comprehensive, cloud-based data engineering framework to facilitate data ingestion, transformation, storage, and visualization. Our primary tools included:

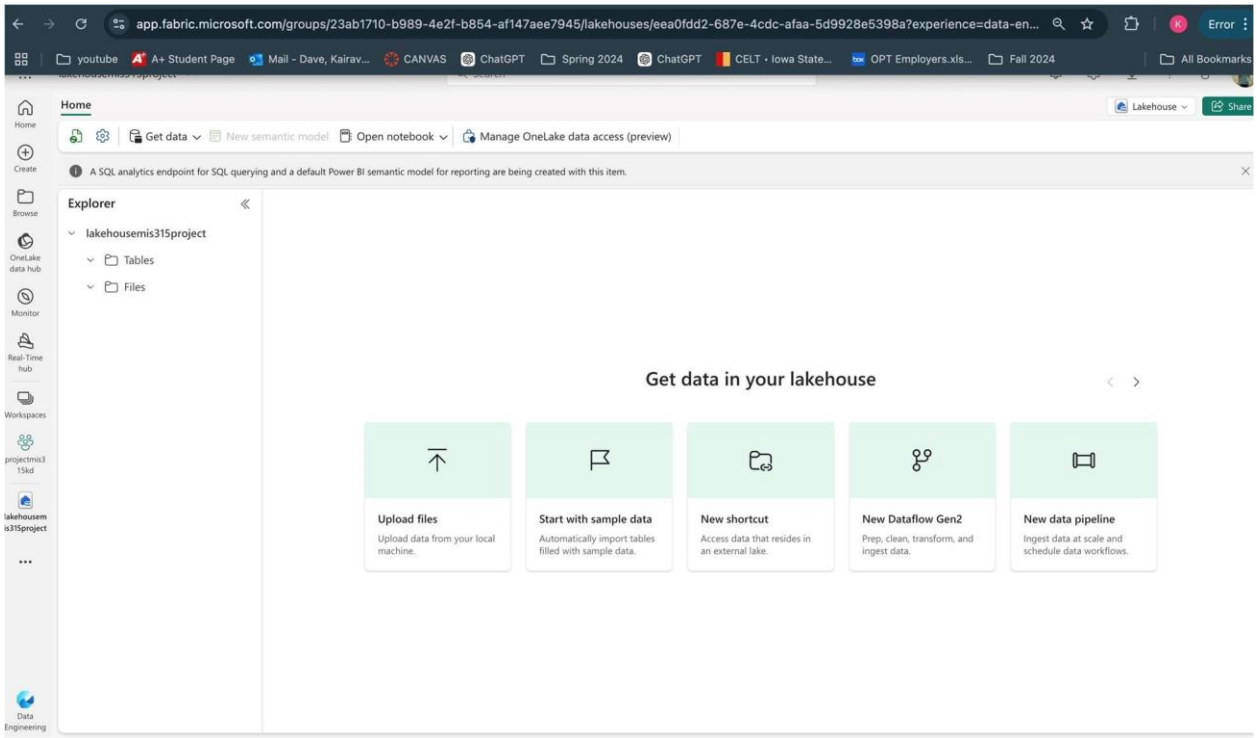
- Azure Synapse Engineering: Enabled data engineering capabilities, allowing for the integration, transformation, and efficient analysis of large datasets.
- Azure Data Factory (ADF): Functioned as the primary tool for data ingestion and pipeline management. Within ADF, we employed Power Query Online to conduct initial data transformations in an efficient, low-code environment.
    - Microsoft Fabric Lakehouse: Served as the central repository for transformed data, providing a structured, scalable, and accessible storage solution for subsequent analysis.
- Power BI: Leveraged to visualize data stored in the Lakehouse, enabling the creation of detailed dashboards and reports to present key findings based on the transformed data

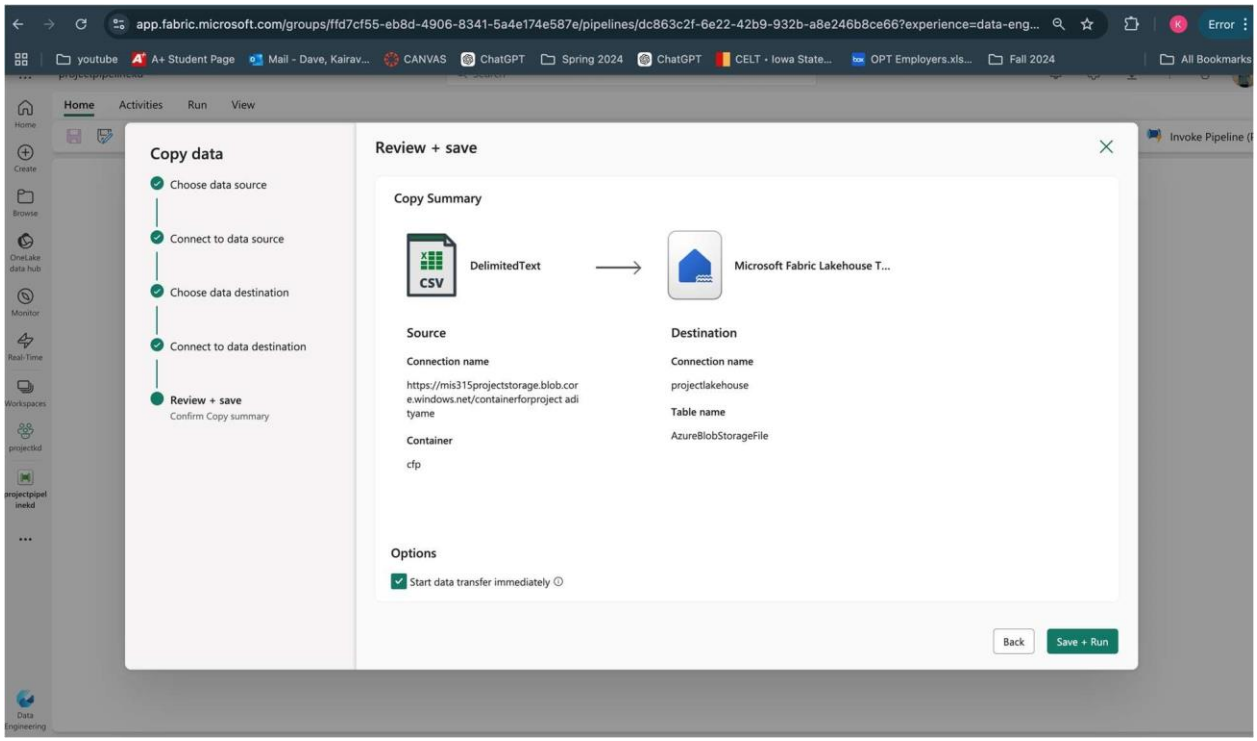
Data Ingestion

- Imported the dataset into Azure Blob Storage to store and facilitate cloud-based access, allowing for scalable processing.



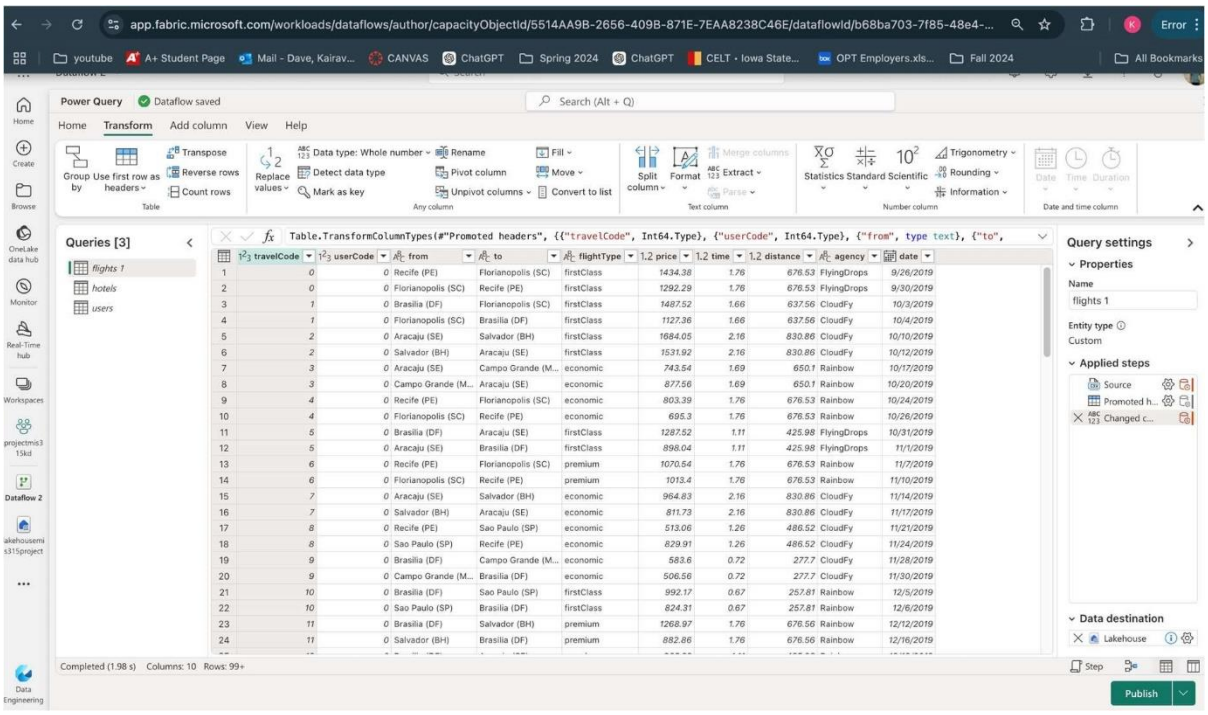
Created an ADF pipeline to transfer data from Blob Storage to Microsoft Fabric Lakehouse for processing and storage.





Data Transformation in Azure Data Factory (Low-Code Dataflows)

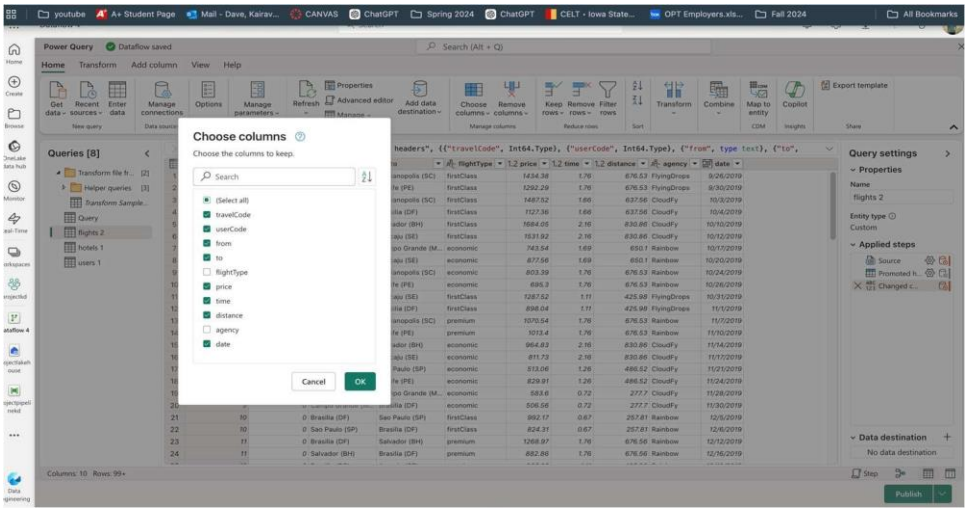
- Set up Dataflows in ADF for data cleaning, filtering, and preparation:

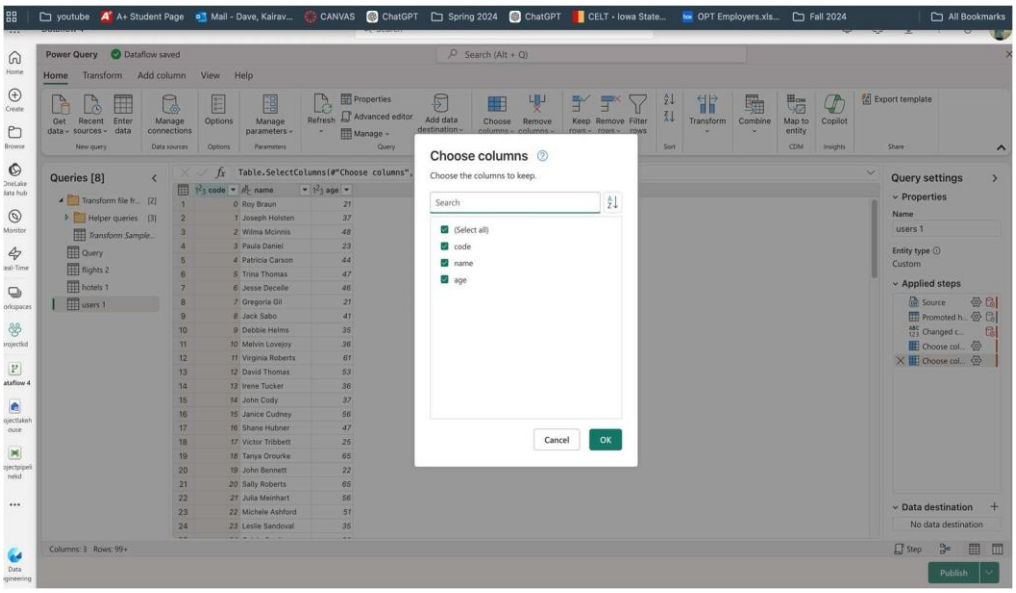


- Loaded the three datasets from the Get data option following Azure Blob Storage and set our data destination as Lakehouse

Column Selection: Selected relevant columns for each dataset.

- Flights: TravelCode, userCode, From, To, Price, Time, Distance, Date
- Hotels: TravelCode, userCode, Name, Place, Days, Price, Total, Date
- Users: Code, Name, Age



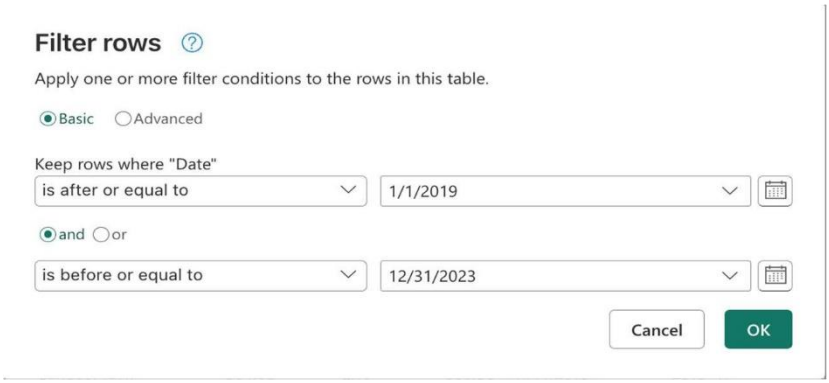


Data Transformation in Power Query Online: Using Power Query Online within ADF, we performed key transformations to prepare the data for analysis:

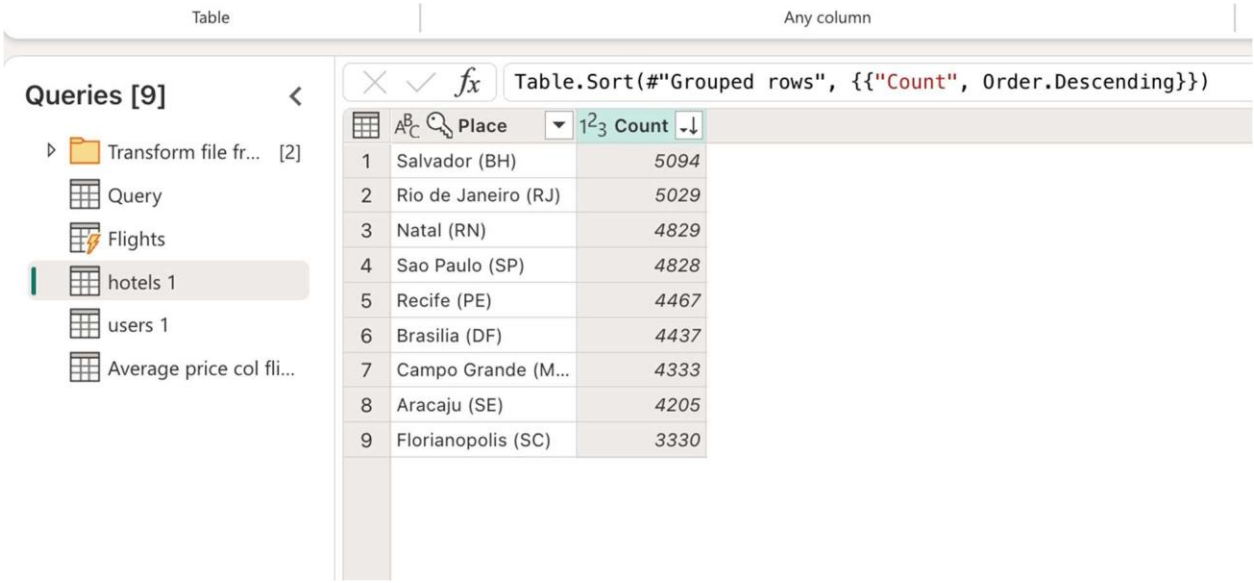
Filtering Data

To focus on recent, relevant data, we applied the following filters:

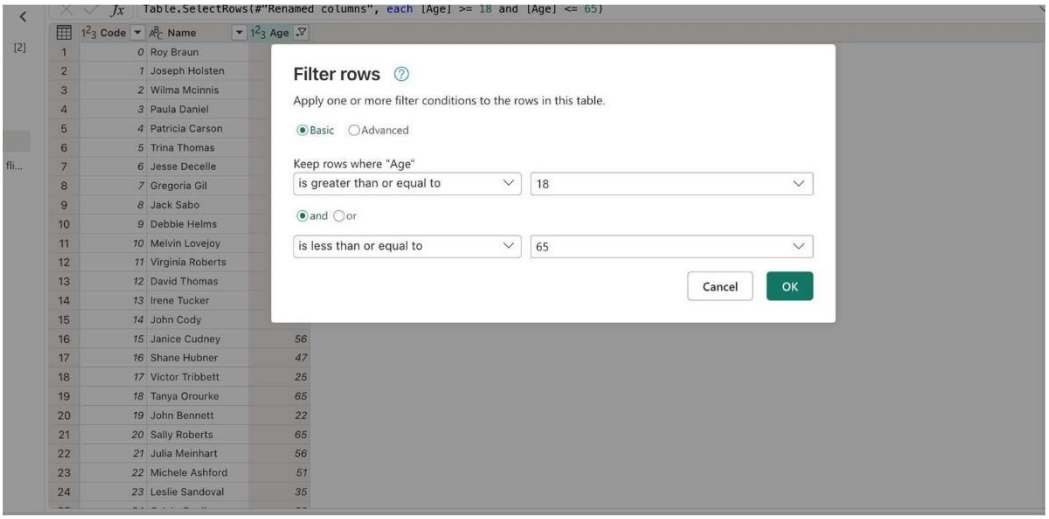
- Date Range: Limited the Flights and Hotels datasets to records from 2019 to 2023, ensuring our analysis reflected recent travel patterns.



- Popular Destinations: Filtered the Hotels dataset to focus on popular destinations



- Age and Trip Frequency: In the Users dataset, we filtered for travelers aged 18-65 and those with multiple trips, focusing our analysis on active, engaged travelers.

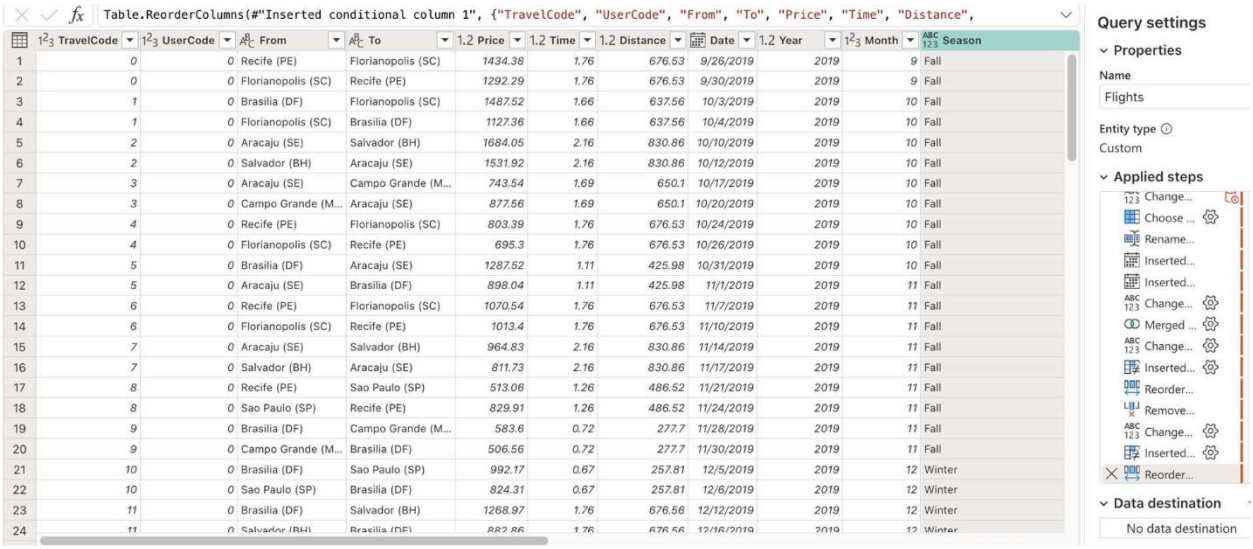


We enhanced the datasets by creating new columns tailored to our analysis goals:

- Year and Month: From the Date column in the Flights and Hotels datasets, we extracted the Year and Month to analyze trends by time period.

Date	1.2 Year	Month
9/26/2019	2019	9
9/30/2019	2019	9
10/3/2019	2019	10
10/4/2019	2019	10
10/10/2019	2019	10
10/12/2019	2019	10
10/17/2019	2019	10
10/20/2019	2019	10
10/24/2019	2019	10
10/26/2019	2019	10
10/31/2019	2019	10
11/1/2019	2019	11

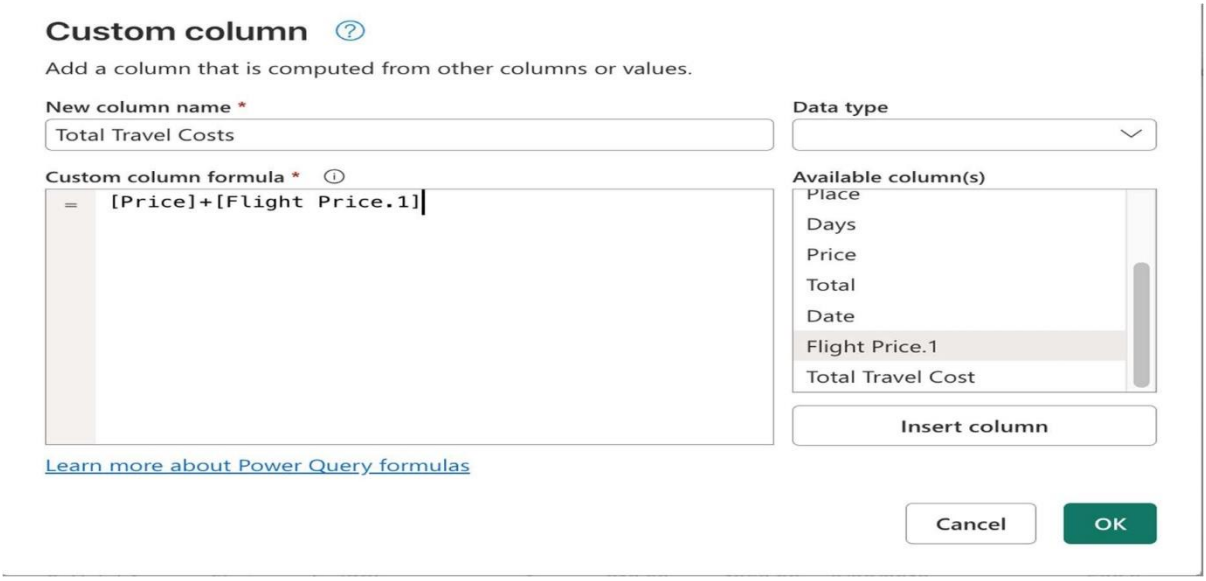
- Season: Based on the Month column, we categorized each record in Flights and Hotels into Winter, Spring, Summer, or Fall, allowing us to analyze seasonal travel patterns.

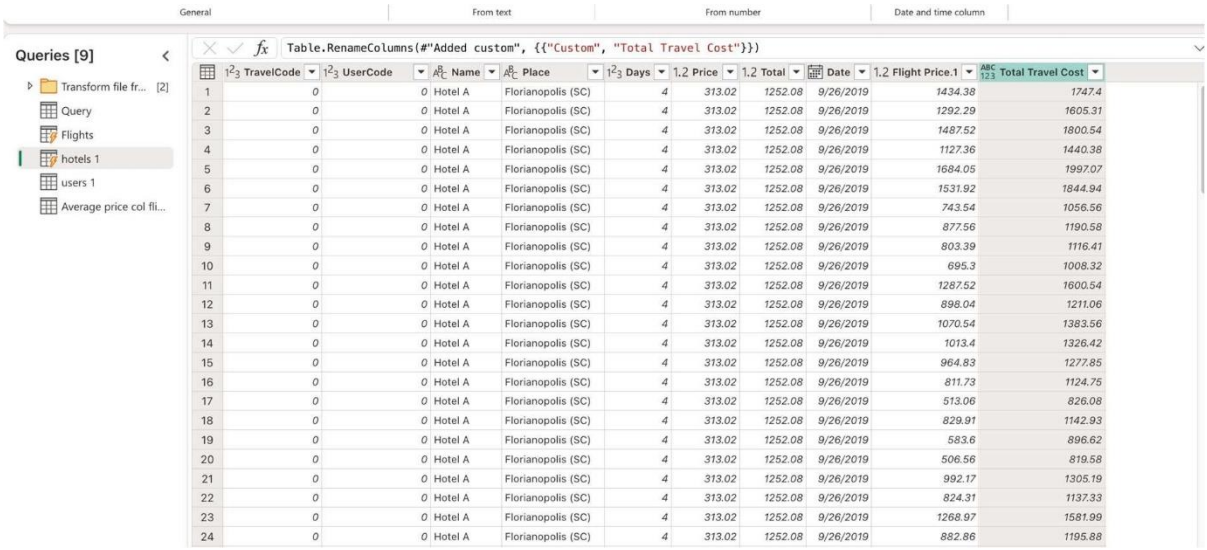


	TravelCode	UserCode	From	To	Price	Time	Distance	Date	Year	Month	Season
1	0	0	Recife (PE)	Florianopolis (SC)	1434.38	1.76	676.53	9/26/2019	2019	9	Fall
2	0	0	Florianopolis (SC)	Recife (PE)	1292.29	1.76	676.53	9/30/2019	2019	9	Fall
3	1	0	Brasilia (DF)	Florianopolis (SC)	1487.52	1.66	637.56	10/3/2019	2019	10	Fall
4	1	0	Florianopolis (SC)	Brasilia (DF)	1127.36	1.66	637.56	10/4/2019	2019	10	Fall
5	2	0	Aracaju (SE)	Salvador (BH)	1684.05	2.16	830.86	10/10/2019	2019	10	Fall
6	2	0	Salvador (BH)	Aracaju (SE)	1531.92	2.16	830.86	10/12/2019	2019	10	Fall
7	3	0	Aracaju (SE)	Campo Grande (M...	743.54	1.69	650.1	10/17/2019	2019	10	Fall
8	3	0	Campo Grande (M...	Aracaju (SE)	877.56	1.69	650.1	10/20/2019	2019	10	Fall
9	4	0	Recife (PE)	Florianopolis (SC)	803.39	1.76	676.53	10/24/2019	2019	10	Fall
10	4	0	Florianopolis (SC)	Recife (PE)	695.3	1.76	676.53	10/26/2019	2019	10	Fall
11	5	0	Brasilia (DF)	Aracaju (SE)	1287.52	1.11	425.98	10/31/2019	2019	10	Fall
12	5	0	Aracaju (SE)	Brasilia (DF)	898.04	1.11	425.98	11/1/2019	2019	11	Fall
13	6	0	Recife (PE)	Florianopolis (SC)	1070.54	1.76	676.53	11/7/2019	2019	11	Fall
14	6	0	Florianopolis (SC)	Recife (PE)	1013.4	1.76	676.53	11/10/2019	2019	11	Fall
15	7	0	Aracaju (SE)	Salvador (BH)	964.83	2.16	830.86	11/14/2019	2019	11	Fall
16	7	0	Salvador (BH)	Aracaju (SE)	811.73	2.16	830.86	11/17/2019	2019	11	Fall
17	8	0	Recife (PE)	Sao Paulo (SP)	513.06	1.26	486.52	11/21/2019	2019	11	Fall
18	8	0	Sao Paulo (SP)	Recife (PE)	829.91	1.26	486.52	11/24/2019	2019	11	Fall
19	9	0	Brasilia (DF)	Campo Grande (M...	583.6	0.72	277.7	11/28/2019	2019	11	Fall
20	9	0	Campo Grande (M...	Brasilia (DF)	506.56	0.72	277.7	11/30/2019	2019	11	Fall
21	10	0	Brasilia (DF)	Sao Paulo (SP)	992.17	0.67	257.81	12/5/2019	2019	12	Winter
22	10	0	Sao Paulo (SP)	Brasilia (DF)	824.31	0.67	257.81	12/6/2019	2019	12	Winter
23	11	0	Brasilia (DF)	Salvador (BH)	1268.97	1.76	676.56	12/12/2019	2019	12	Winter
24	11	0	Salvador (BH)	Brasilia (DF)	882.86	1.76	676.56	12/16/2019	2019	12	Winter

- Total Travel Cost: Calculated by summing flight and hotel expenses per user, this metric allowed us to assess the impact of travel budgets on destination choices.

  We merged the Hotel and Flights by UserCode and added a custom column which calculates the Total Travel Cost



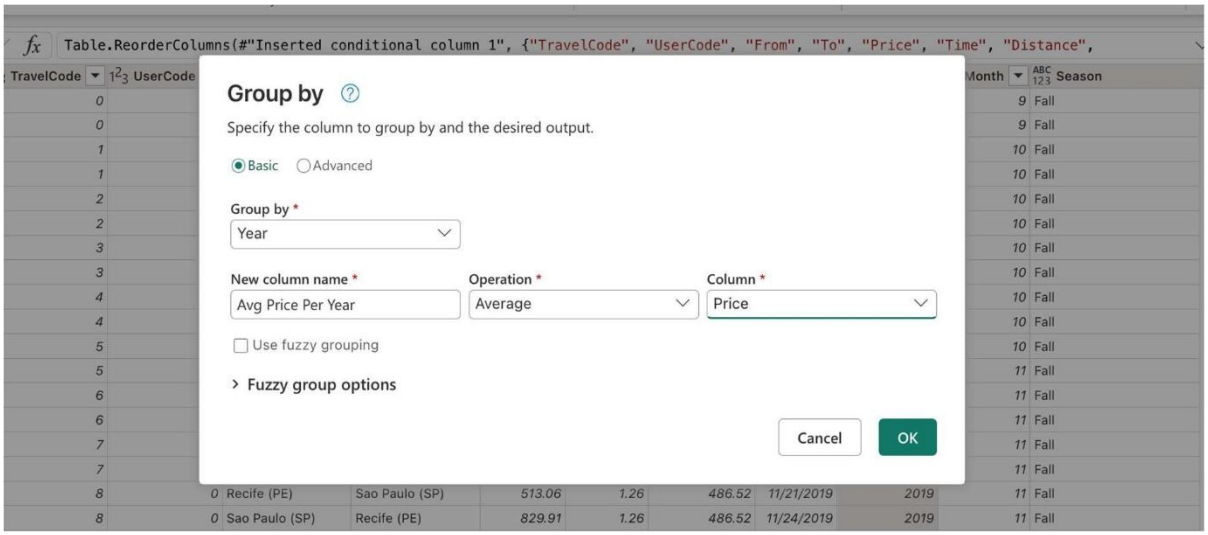


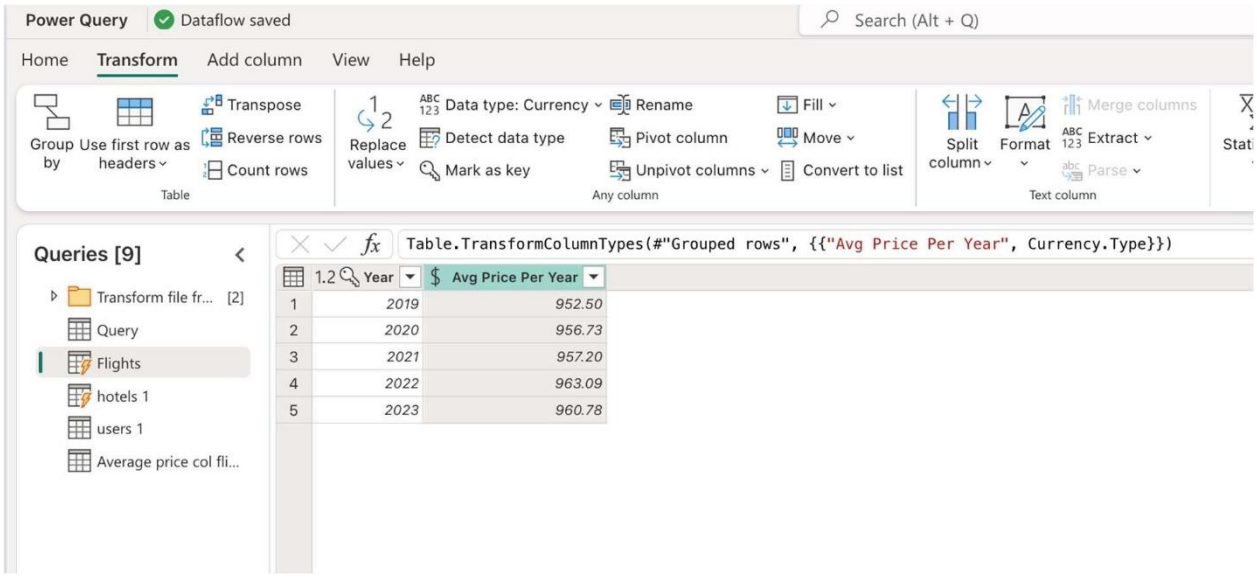
	TravelCode	UserCode	Name	Place	Days	Price	Total	Date	Flight Price.1	Total Travel Cost
1	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1434.38	1747.4
2	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1292.29	1605.31
3	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1487.52	1800.54
4	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1127.36	1440.38
5	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1684.05	1997.07
6	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1531.92	1844.94
7	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	743.54	1056.56
8	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	877.56	1190.58
9	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	803.39	1116.41
10	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	695.3	1008.32
11	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1287.52	1600.54
12	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	898.04	1211.06
13	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1070.54	1383.56
14	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1013.4	1326.42
15	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	964.83	1277.85
16	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	811.73	1124.75
17	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	513.06	826.08
18	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	829.91	1142.93
19	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	583.6	896.62
20	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	506.56	819.58
21	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	992.17	1305.19
22	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	824.31	1137.33
23	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	1268.97	1581.99
24	0	0	Hotel A	Florianopolis (SC)	4	313.02	1252.08	9/26/2019	882.86	1195.88

Data Aggregation

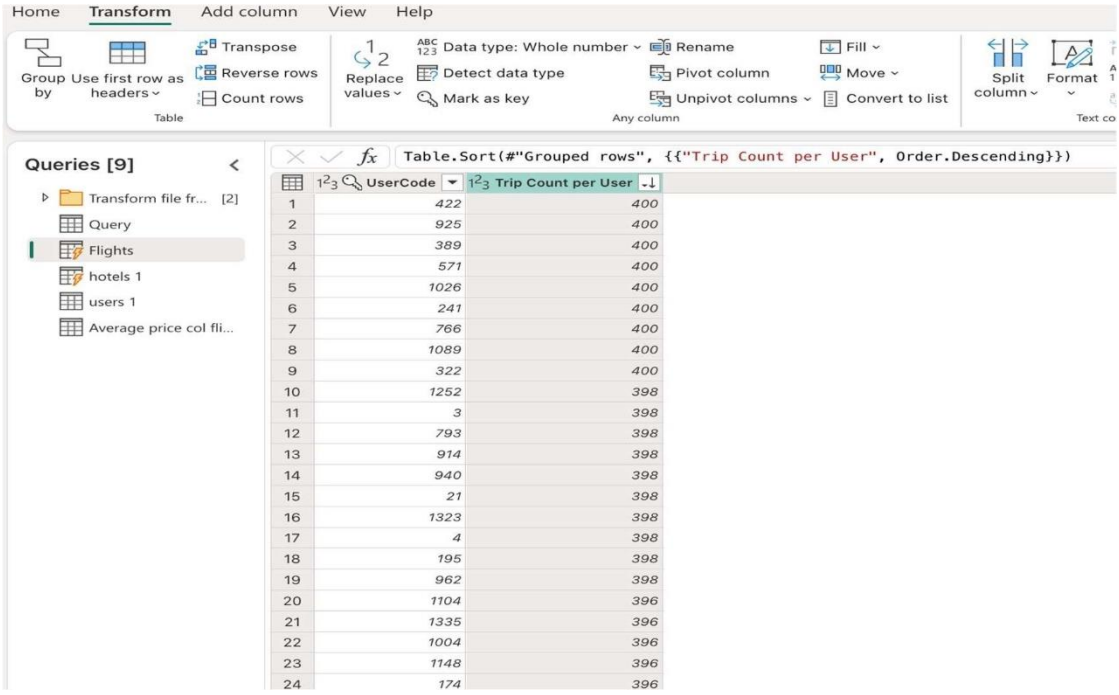
We applied aggregations to capture essential metrics for analysis:

- Average Price by Year: In the Flights dataset, we calculated the average flight price by year, which helped us track changes in travel costs over time.

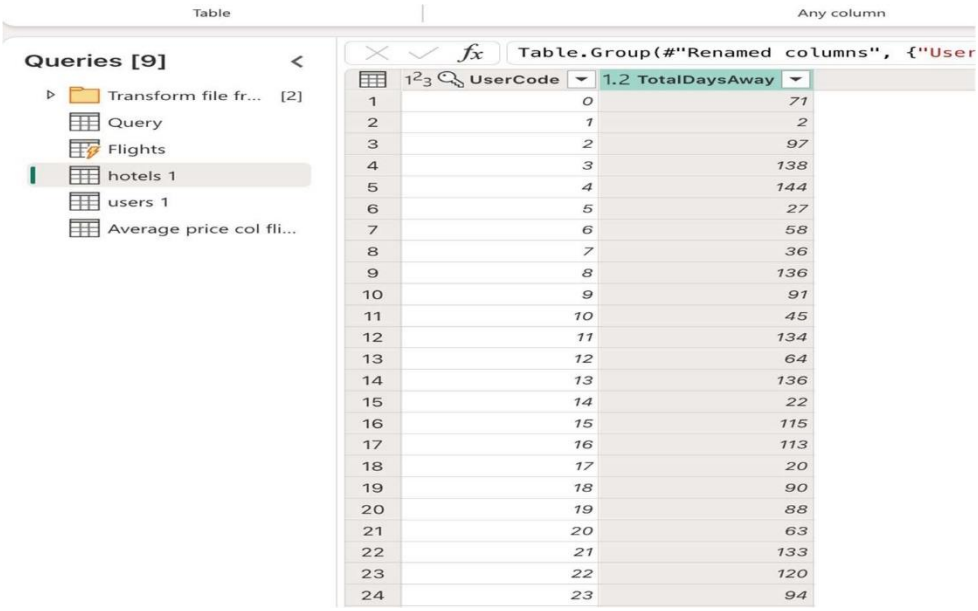




- Trip Count per User: Aggregated the total number of trips per user in the Users dataset, allowing us to identify frequent travelers.



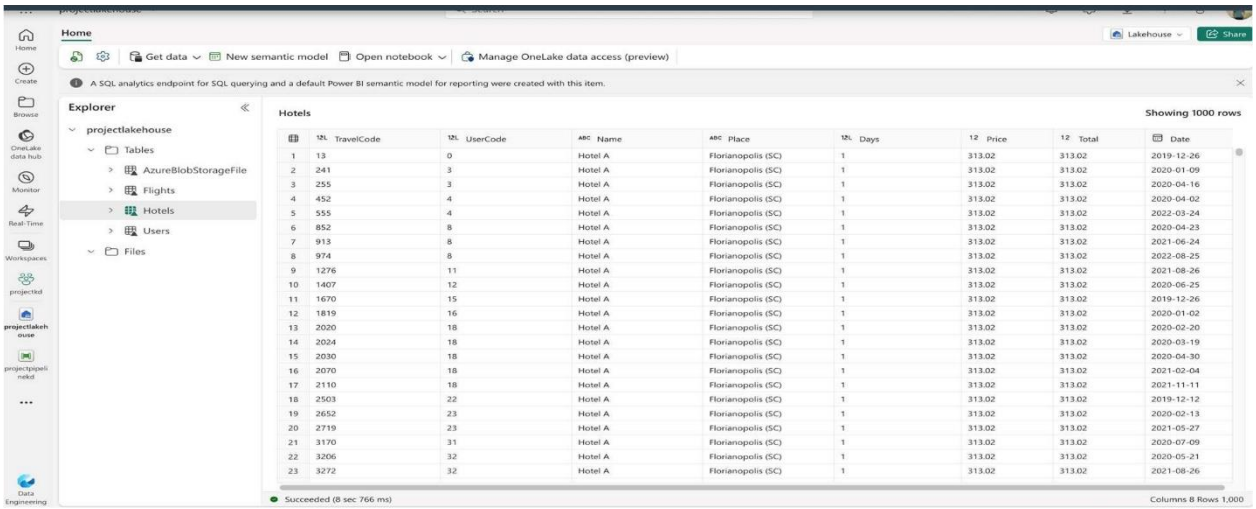
- Total Days Away: For each customer, we summed up the duration of all hotels stays, helping us understand extended travel patterns.



	UserCode	TotalDaysAway
1	0	71
2	1	2
3	2	97
4	3	138
5	4	144
6	5	27
7	6	58
8	7	36
9	8	136
10	9	91
11	10	45
12	11	134
13	12	64
14	13	136
15	14	22
16	15	115
17	16	113
18	17	20
19	18	90
20	19	88
21	20	63
22	21	133
23	22	120
24	23	94

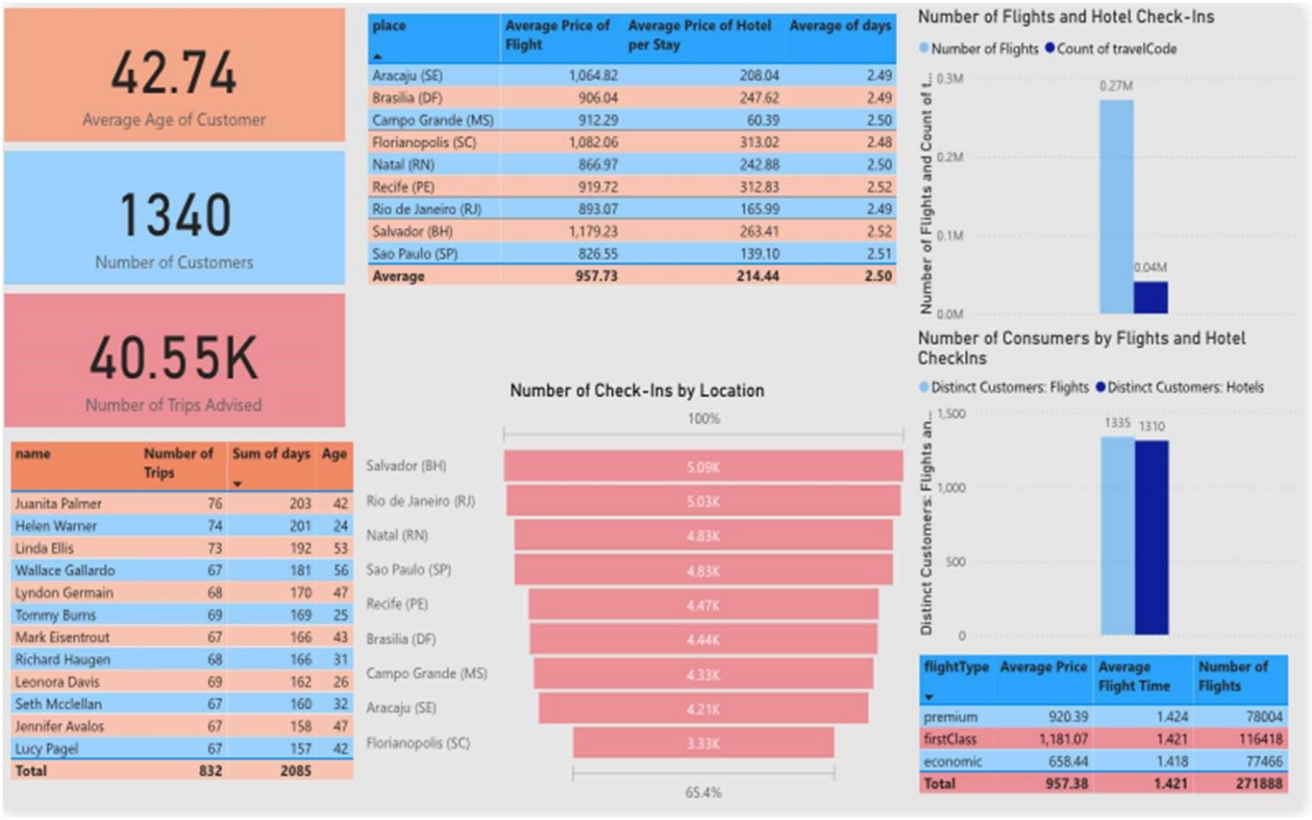
Data Flow

After completing all transformations, we stored the data in Microsoft Fabric Lakehouse, creating an organized and scalable structure that ensured easy accessibility for further analysis and reporting. We then connected the transformed Lakehouse data to a pipeline to have our refined datasets stored in our Container in Storage Account in azure and uploaded in Power BI, where we developed interactive visualizations and dashboards to effectively derive insights and present our findings in a clear, impactful manner.



Power-BI Descriptive Analytics Dashboard

This Power BI dashboard provides descriptive analytics for customer travel data, offering insights into customer demographics, travel patterns, and pricing. Key metrics such as average age, number of customers, average price of flights and hotels, and check-in frequencies across various locations help highlight trends and customer preferences, supporting data-driven decision-making for travel planning and optimization.



place	Average Price of Flight	Average Price of Hotel per Stay	Average of days
Aracaju (SE)	1,064.82	208.04	2.49
Brasilia (DF)	906.04	247.62	2.49
Campo Grande (MS)	912.29	60.39	2.50
Florianopolis (SC)	1,082.06	313.02	2.48
Natal (RN)	866.97	242.88	2.50
Recife (PE)	919.72	312.83	2.52
Rio de Janeiro (RJ)	893.07	165.99	2.49
Salvador (BH)	1,179.23	263.41	2.52
Sao Paulo (SP)	826.55	139.10	2.51
Average	**957.73**	**214.44**	**2.50**

name	Number of Trips	Sum of days	Age
Juanita Palmer	76	203	42
Helen Warner	74	201	24
Linda Ellis	73	192	53
Wallace Gallardo	67	181	56
Lyndon Germain	68	170	47
Tommy Burns	69	169	25
Mark Eisentrout	67	166	43
Richard Haugen	68	166	31
Leonora Davis	69	162	26
Seth Mcclellan	67	160	32
Jennifer Avalos	67	158	47
Lucy Pagel	67	157	42
Total	**832**	**2085**	

flightType	Average Price	Average Flight Time	Number of Flights
premium	920.39	1.424	78004
firstClass	1,181.07	1.421	116418
economic	658.44	1.418	77466
Total	**957.38**	**1.421**	**271888**

Image 1: Interactive Dashboard displaying information regarding users, flights, and hotels.

Image 2: Average age, number of customers, and number of trips (by hotels) advised

place	Average Price of Flight	Average Price of Hotel per Stay	Average of days
Aracaju (SE)	1,064.82	208.04	2.49
Brasilia (DF)	906.04	247.62	2.49
Campo Grande (MS)	912.29	60.39	2.50
Florianopolis (SC)	1,082.06	313.02	2.48
Natal (RN)	866.97	242.88	2.50
Recife (PE)	919.72	312.83	2.52
Rio de Janeiro (RJ)	893.07	165.99	2.49
Salvador (BH)	1,179.23	263.41	2.52
Sao Paulo (SP)	826.55	139.10	2.51
Average	**957.73**	**214.44**	**2.50**

Image 3: By location, average price of flight, average price of hotel per stay, and average length of stay

name	Number of Trips	Sum of days	Age
Juanita Palmer	76	203	42
Helen Warner	74	201	24
Linda Ellis	73	192	53
Wallace Gallardo	67	181	56
Lyndon Germain	68	170	47
Tommy Burns	69	169	25
Mark Eisentrout	67	166	43
Richard Haugen	68	166	31
Leonora Davis	69	162	26
Seth Mcclellan	67	160	32
Jennifer Avalos	67	158	47
Lucy Pagel	67	157	42
Total	**832**	**2085**	

Image 4: Top 12 customers, number of hotel trips advised, total days per customer spent away, and age

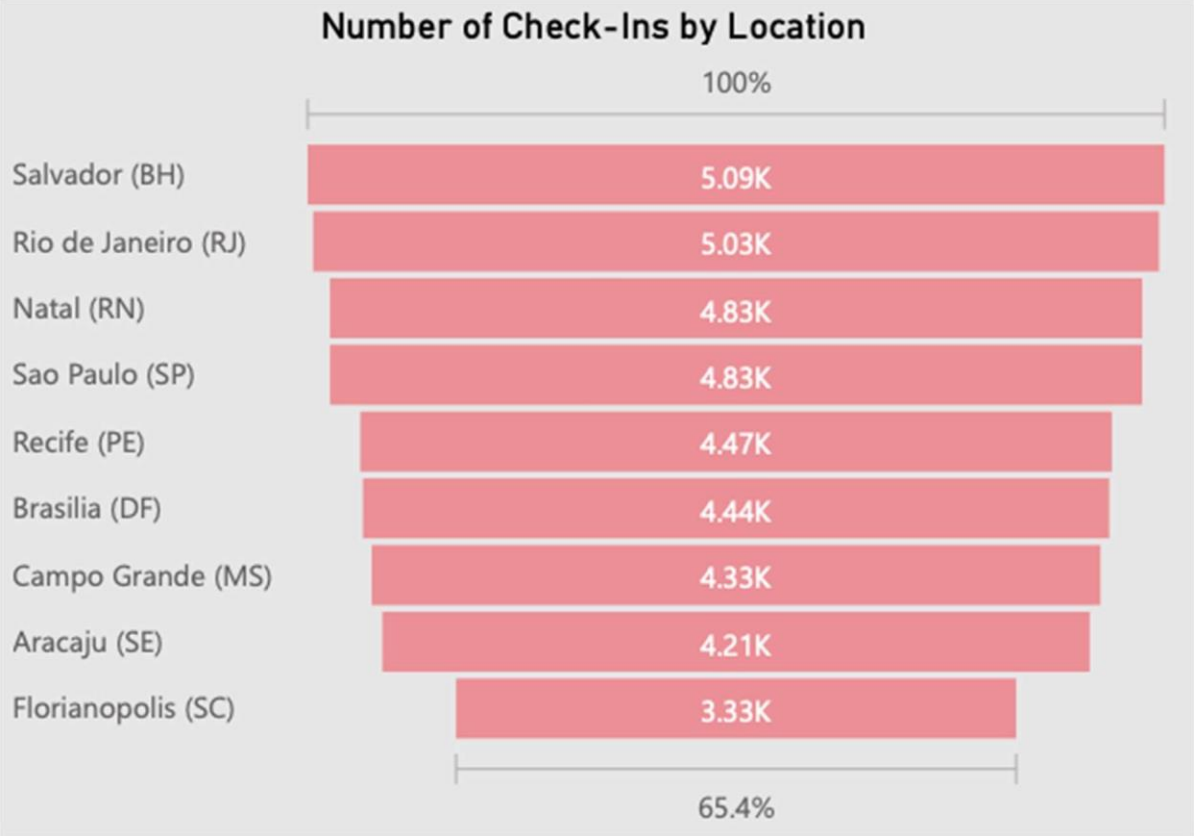


Image 5: Check-ins by location

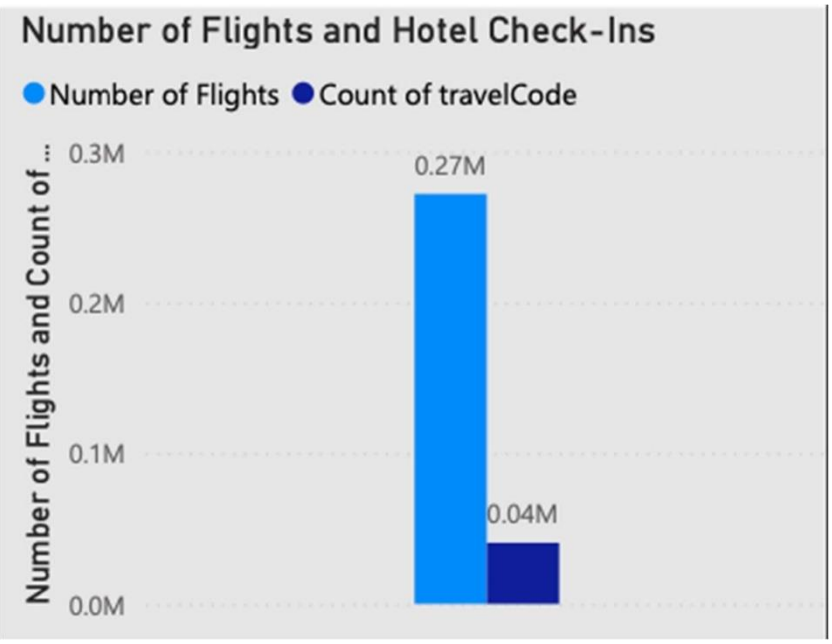


Image 6: Total flights booked, and total hotels trips reserved (Note: Due to some customers booking one-way flights, round-trip flights are counted as two flights).

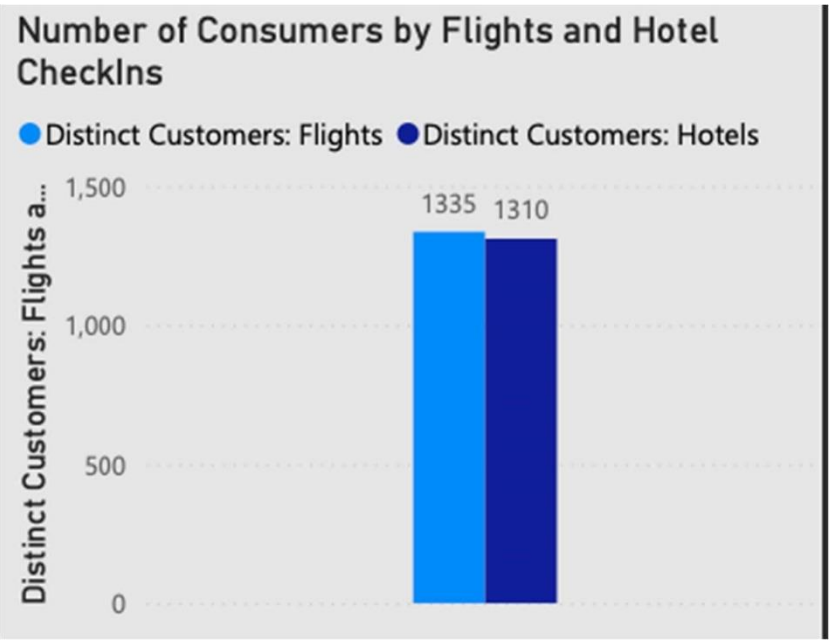


Image 7: Number of customers who booked flights with us and number of customers who booked hotel rooms with us.

flightType	Average Price	Average Flight Time	Number of Flights
premium	920.39	1.424	78004
firstClass	1,181.07	1.421	116418
economic	658.44	1.418	77466
Total	**957.38**	**1.421**	**271888**

Image 8: Summary of flights broken down into class, average price, average flight time, and number of flights booked.

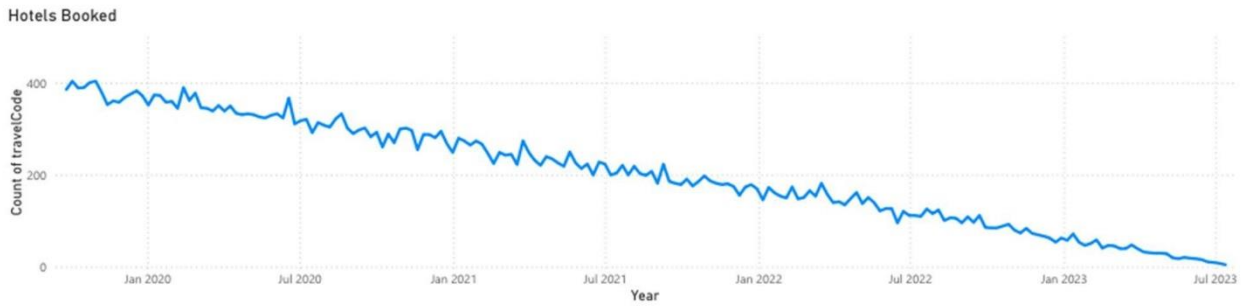


Image 9: Number of hotel check-ins by date from 2019 to 2023

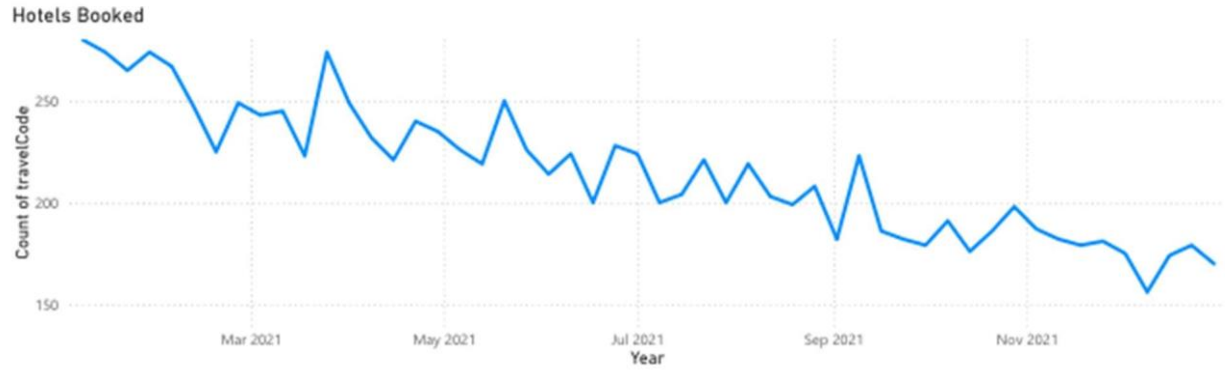


Image 10: A closer look at the number of hotel check-ins for the year 2021.

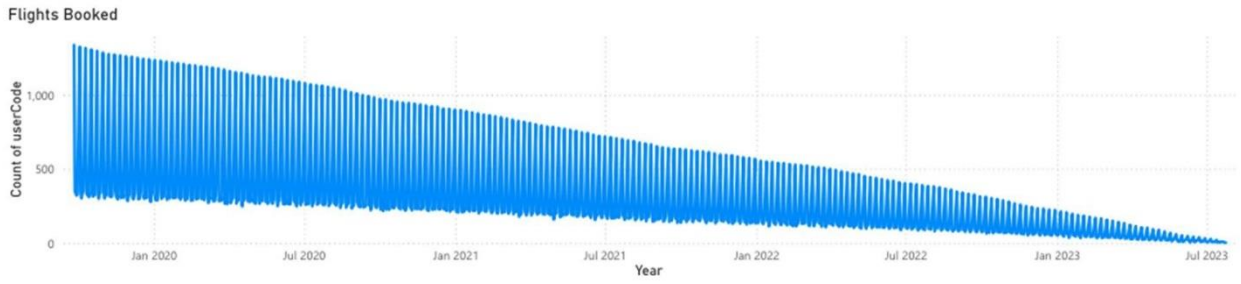


Image 11: Flights booked from 2019-2023. We can see it creates an oscillating pattern and in fact does not indicate any seasonality trends as we previously assumed.

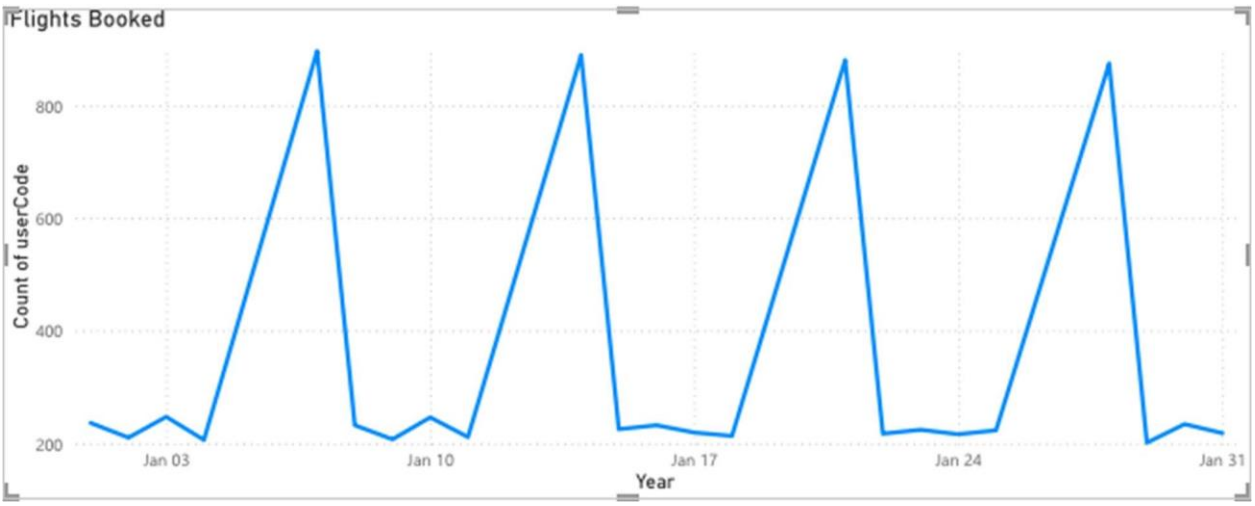
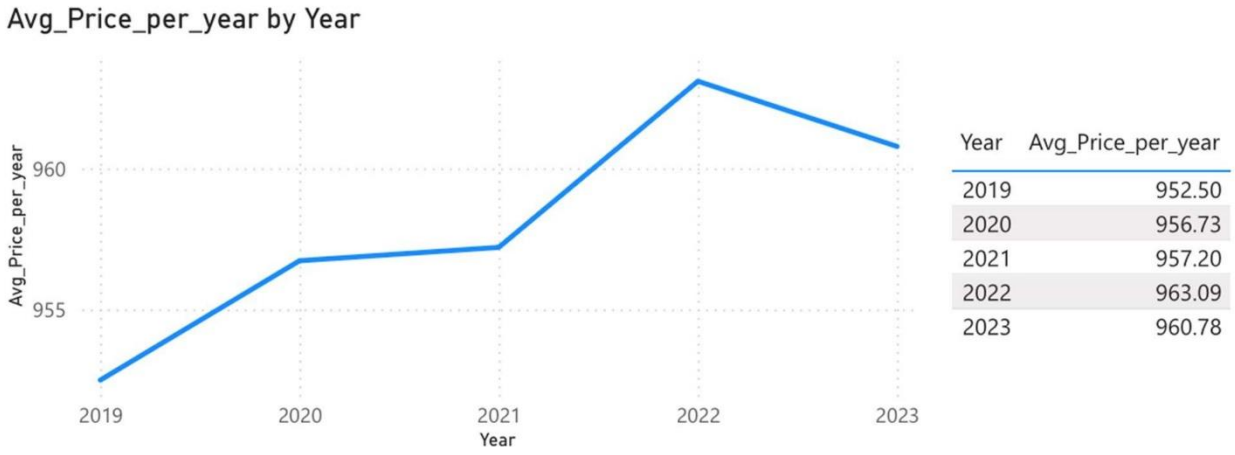


Image 12: A closer look at January 2021 flights. An increase in flights on Thursdays, and a small increase on Sundays.



Year	Avg_Price_per_year
2019	952.50
2020	956.73
2021	957.20
2022	963.09
2023	960.78

Image 13: The average price of flights per year. As expected, there is an upward trend over the past several years.

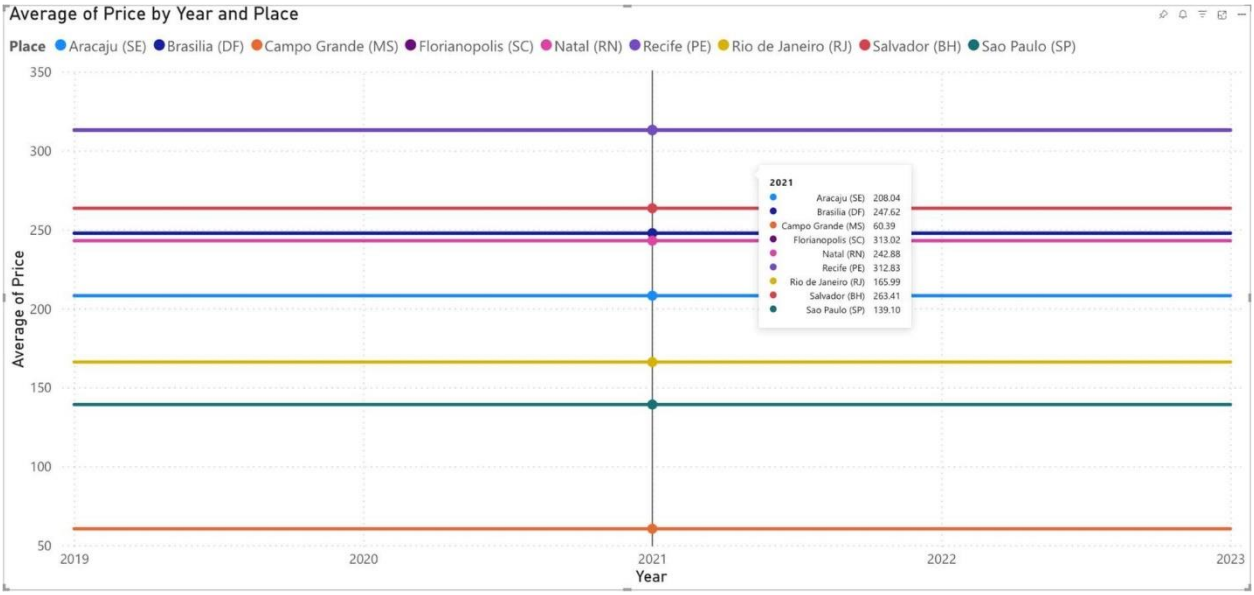


Image 14: The average price of hotels over the per year. The average price per hotel has not changed over the last few years, unlike flights.

Image 15: Summary statistics of how many days people tend to book their trips for.

Recommendations

Customer Demographics

An analysis of customer demographics shows an average age of around 43. As seen in image 4, the top customers are all roughly in the age range of 25 ≈ 50. Considering this age range of top customers and the overall average age of customers, we can identify a target consumer group.

Recommendation:

- We recommend targeting marketing campaigns and customer reward programs at consumers within the ages of 25 - 50 with a focus on the 35 – 45 age range. Based on our

analysis this group of consumers is most likely to fly more often. Therefore, strong customer loyalty within this demographic will be most beneficial.

Time Trends

Flight patterns vary significantly according to time of year. As shown in image 12, the number of flights booked increases significantly over the weekends. This trend is consistent throughout each year within our dataset.

Recommendations:

- To accommodate the increased demand for flights over the weekends, more flight options should be available. This will ensure that the company can keep up with demand and customers do not need to stress about the availability of their preferred flight.
- We also recommend focusing marketing campaigns to increase customer engagement in times of less demand. Reducing flight costs to push flight bookings during the week would encourage consumers to purchase flights at these times.

Customer Loyalty

Recommendations:

- Customer loyalty is very important because according to our data analysis, customers will make many bookings over time. To encourage customers to continue booking with our company, we should offer a customer loyalty program. Returning business will be good for the company, and loyalty rewards will help loyal customers save money.
- We recommend running campaigns to increase customer activity during seasons of lower average bookings. Customer loyalty programs could encourage customers to book flights in the off-season.

Location

An analysis of check-ins by location shows that Salvador and Rio de Janeiro are the top locations for check-ins, with Natal and Sao Paulo following close behind.

Recommendation:

- We recommend focusing flights to the top locations by check-in. We want to ensure the business balances demand between flights and hotels.
- The business could also benefit from increased marketing for less popular locations.