

DELIVERABLE - 2

DELIVERABLE - 2

The dataset is centered on airline passenger satisfaction, encompassing a range of attributes such as gender, age, customer type, travel class, and flight distance. It differentiates types of travel into personal and business categories. The dataset offers detailed ratings for various services, including inflight wifi, seat comfort, food and drink, and cleanliness. Additionally, it provides ratings for specific services like inflight entertainment, on-board service, leg room service, baggage handling, check-in service, and inflight service, presenting a comprehensive view of the passenger's travel experience.

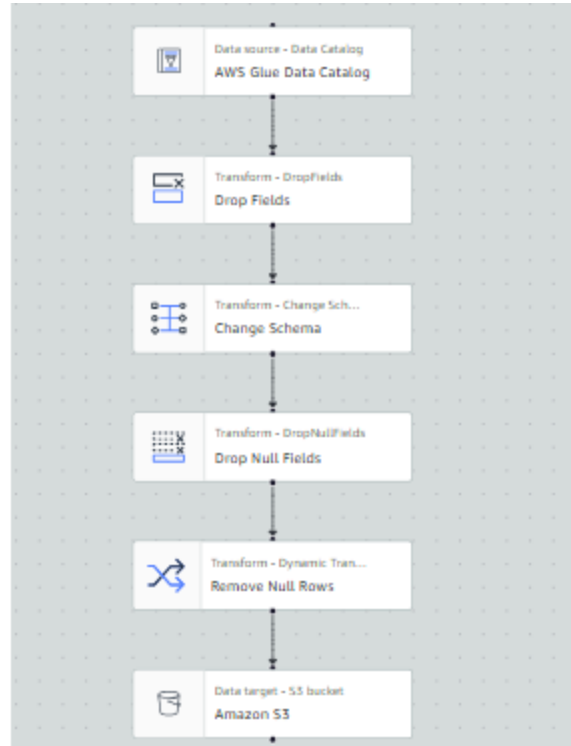
Flight delays are recorded in terms of departure and arrival delay minutes, shedding light on their impact on passenger satisfaction. The overall satisfaction level of passengers is categorized as either 'satisfied' or 'neutral or dissatisfied'. This data is crucial for airlines to analyze and predict factors influencing passenger satisfaction, making it an essential tool for market research, customer experience management, and service quality improvement in the aviation industry. The dataset's comprehensive nature makes it invaluable for understanding passenger preferences and improving airline services.

Data Understanding

We employed AWS Glue for schema discovery and data type definition, ensuring our dataset's structure was accurately recognized and categorized. Then we used Glue crawler for refining the dataset schema, allowing for precise categorization and organization of data fields. For in-depth querying and extracting insights, AWS Athena was used. This enabled us to perform comprehensive data analysis efficiently, allowing for complex queries and data aggregations. This approach not only streamlined our data analysis process but also enhanced the accuracy and relevance of the insights gained, proving crucial for informed decision-making and strategic planning. Additionally, we utilized AWS QuickSight for data visualization. QuickSight's powerful visualization tools provided an intuitive interface for exploring and presenting the data, making it easier to identify patterns, trends, and correlations. This visual approach to data analysis enhanced our ability to gain insights, allowing us to make more informed decisions and develop effective strategies based on the dataset.

DELIVERABLE - 2

AWS GLUE PIPELINE



AWS Glue Schema (23)

View and manage the table schema.

Filter schemas

#	Column name	Data type	Partition key	Comment
1	gender	string	-	-
2	customer_type	string	-	-
3	age	bigint	-	-
4	type_of_travel	string	-	-
5	class	string	-	-
6	flight_distance	bigint	-	-
7	inflight_wifi_service	bigint	-	-
8	departure/arrival_time_c...	bigint	-	-
9	ease_of_online_booking	bigint	-	-
10	gate_location	bigint	-	-
11	food_and_drink	bigint	-	-
12	online_boarding	bigint	-	-
13	seat comfort	bigint	-	-
14	inflight_entertainment	bigint	-	-
15	on_board_service	bigint	-	-
16	leg_room_service	bigint	-	-
17	baggage_handling	bigint	-	-
18	checkin_service	bigint	-	-
19	inflight_service	bigint	-	-

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

DELIVERABLE - 2

AWS ATHENA

The screenshot shows the AWS Athena console interface. The top navigation bar includes the AWS logo, a search bar, and user information. The left sidebar contains a menu with 'Editor', 'Recent queries', 'Saved queries', and 'Settings'. The main area displays the 'Editor' tab for 'Query 14'. The SQL query is:

```
1 SELECT Type_of_Travel, AVG(Flight_Distance) as Average_Distance
2 FROM myairline_output_data
3 GROUP BY Type_of_Travel;
```

The query has been executed successfully, as indicated by the 'Completed' status. The results are shown in a table with 2 rows:

#	Type_of_Travel	Average_Distance
1	"Personal Travel"	792.0810567769544
2	"Business travel"	1368.2872374572605

The console also shows the 'Data' section on the left, listing tables and views, and the 'Query results' section at the bottom.

The screenshot shows the AWS Athena console interface. The top navigation bar includes the AWS logo, a search bar, and user information. The left sidebar contains a menu with 'Editor', 'Recent queries', 'Saved queries', and 'Settings'. The main area displays the 'Editor' tab for 'Query 13'. The SQL query is:

```
1 SELECT Gender, AVG(Inflight_wifi_service) as Average_Wifi_Rating
2 FROM myairline_output_data
3 GROUP BY Gender;
```

The query has been executed successfully, as indicated by the 'Completed' status. The results are shown in a table with 2 rows:

#	Gender	Average_Wifi_Rating
1	Male	2.741778533325517
2	Female	2.7179433686725964

The console also shows the 'Data' section on the left, listing tables and views, and the 'Query results' section at the bottom.

DELIVERABLE - 2

The screenshot shows the AWS Athena console interface. On the left, the 'Data' sidebar is visible with 'Data source' set to 'AwsDataCatalog' and 'Database' set to 'myairline-db'. Below this, a list of tables is shown, including 'departure/arrival_time_convenient', 'ease_of_online_booking', 'gate_location', 'food_and_drink', 'online_boarding', 'seat comfort', 'inflight_entertainment', 'on_board_service', 'leg_room_service', 'baggage_handling', and 'checkin_service'. The main panel displays 'Query 11' with the following SQL query:

```
1 SELECT satisfaction, MAX(departure_delay_in_minutes) as Max_Departure_Delay
2 FROM myairline_output_data
3 GROUP BY satisfaction;
```

The query status is 'Completed' with a time in queue of 142 ms, run time of 708 ms, and data scanned of 10.86 MB. The results table shows 2 rows:

#	satisfaction	Max_Departure_Delay
1	"neutral or dissatisfied"	1592
2	satisfied	1305

The screenshot shows the AWS Athena console interface. On the left, the 'Data' sidebar is visible with 'Data source' set to 'AwsDataCatalog' and 'Database' set to 'myairline-db'. Below this, a list of tables is shown, including 'departure/arrival_time_convenient', 'ease_of_online_booking', 'gate_location', 'food_and_drink', 'online_boarding', 'seat comfort', 'inflight_entertainment', 'on_board_service', 'leg_room_service', 'baggage_handling', 'checkin_service', and 'inflight_service'. The main panel displays 'Query 12' with the following SQL query:

```
1 SELECT Class, satisfaction, COUNT(*) as Count
2 FROM myairline_output_data
3 GROUP BY Class, satisfaction;
```

The query status is 'Completed' with a time in queue of 106 ms, run time of 908 ms, and data scanned of 10.86 MB. The results table shows 6 rows:

#	Class	satisfaction	Count
1	"Eco Plus"	"neutral or dissatisfied"	5650
2	Business	satisfied	34480
3	Eco	satisfied	8701
4	Business	"neutral or dissatisfied"	15185
5	Eco	"neutral or dissatisfied"	38044
6	"Eco Plus"	satisfied	1844

DELIVERABLE - 2

The screenshot displays the AWS Athena console interface. The top navigation bar shows the AWS logo, a search bar, and various service icons (VPC, EC2, CloudWatch, Amazon SageMaker). The main header includes tabs for 'Editor', 'Recent queries', 'Saved queries', and 'Settings', along with a 'Workgroup' dropdown set to 'primary'.

The left sidebar, titled 'Data', contains sections for 'Data source' (set to 'AwsDataCatalog'), 'Database' (set to 'myairline-db'), and 'Tables and views'. Under 'Tables (2)', 'myairline_input_data' and 'myairline_output_data' are listed. Under 'Views (0)', no views are shown.

The main editor area shows a SQL query for 'Query 10':

```
1 SELECT Customer_Type, AVG(Age) as Average_Age
2 FROM myairline_output_data
3 GROUP BY Customer_Type;
```

Below the query editor, there are buttons for 'Run again', 'Explain', 'Cancel', 'Clear', and 'Create'. A 'Reuse query results' toggle is also present, indicating results are available up to 60 minutes ago.

The 'Query results' section shows a status of 'Completed' with performance metrics: 'Time in queue: 102 ms', 'Run time: 736 ms', and 'Data scanned: 10.86 MB'. Buttons for 'Copy' and 'Download results' are available.

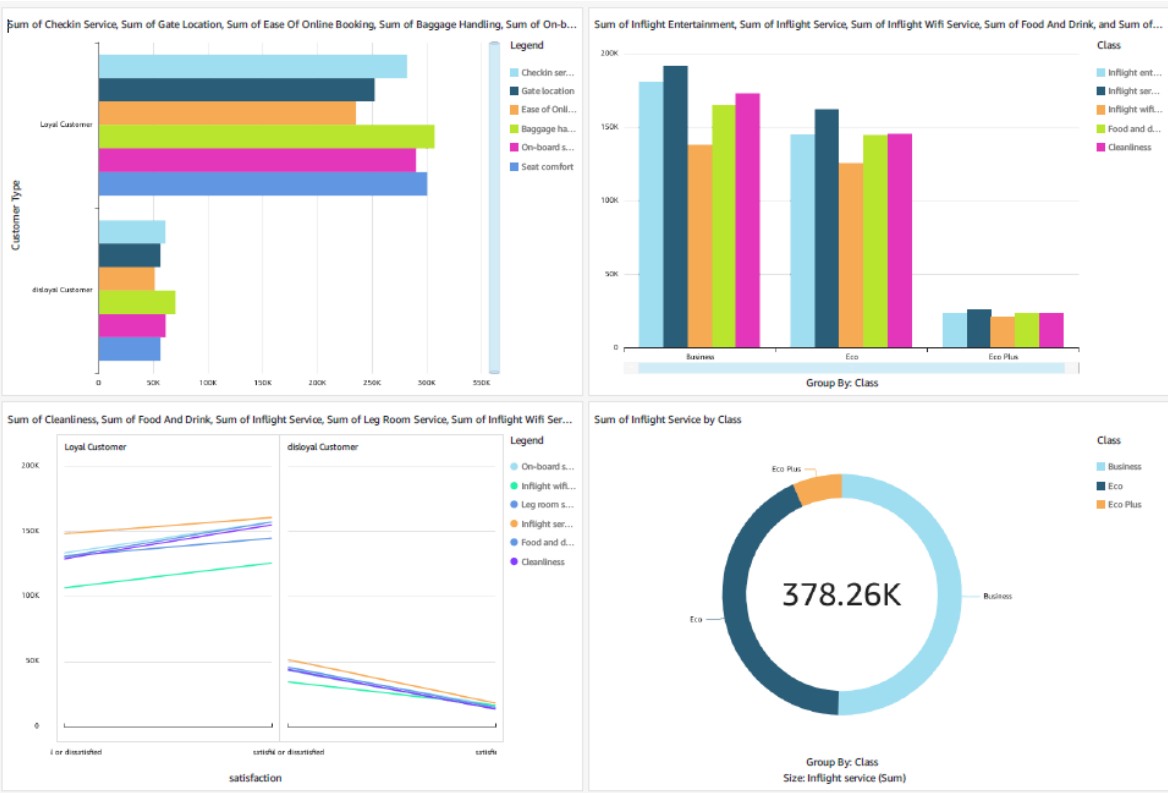
The 'Results (2)' section displays a table with the following data:

#	Customer_Type	Average_Age
1	"Loyal Customer"	41.392496732333996
2	"disloyal Customer"	30.374269005847953

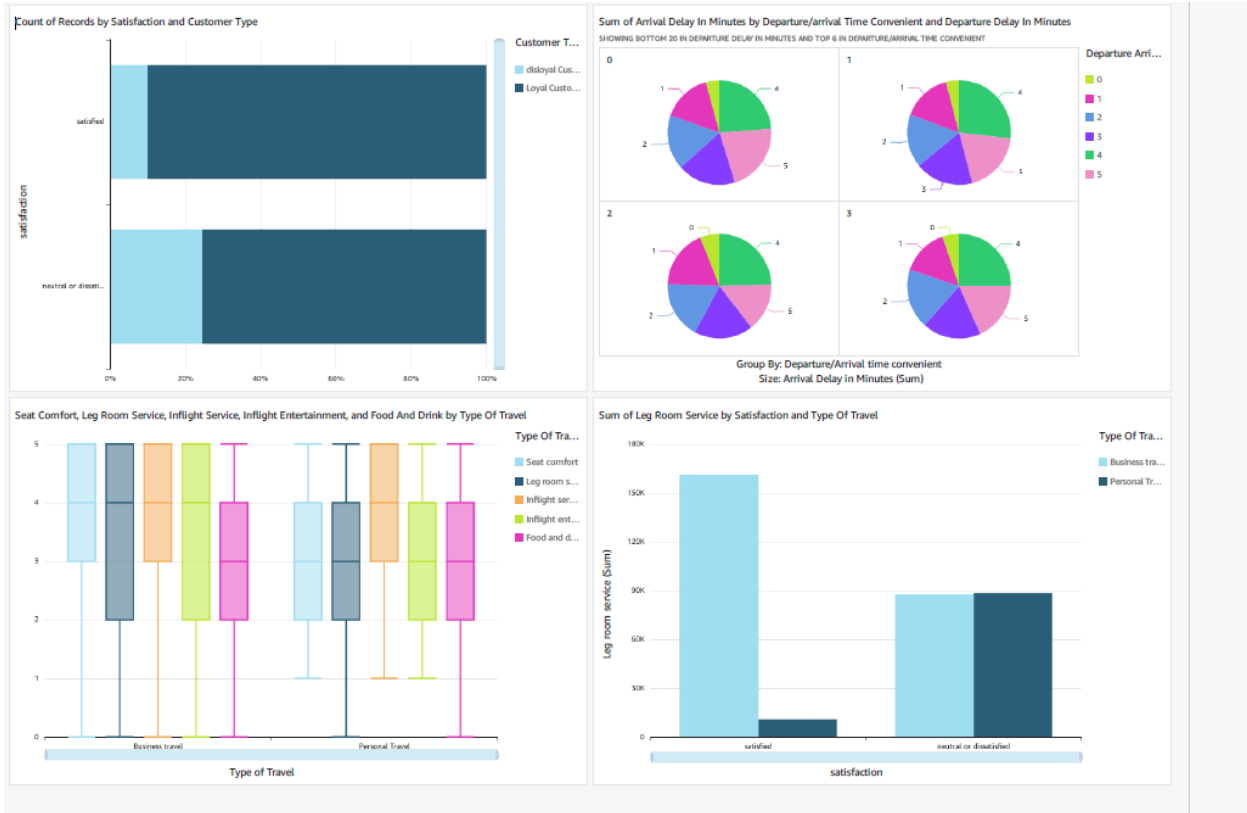
The bottom of the console features a footer with 'CloudShell', 'Feedback', and copyright information for Amazon Web Services, Inc. or its affiliates, along with links for 'Privacy', 'Terms', and 'Cookie preferences'.

DELIVERABLE - 2

AWS QUICKSITE



DELIVERABLE - 2



AWS S3 BUCKET

s3.console.aws.amazon.com/s3/buckets?region=us-east-1

Services Search [Option+S] Global voclabs/user2743281=savasar1@uncc.edu @ 5190-0554-2398

VPC EC2 CloudWatch Amazon SageMaker

Amazon S3

- Buckets
 - Access Points
 - Object Lambda Access Points
 - Multi-Region Access Points
 - Batch Operations
 - IAM Access Analyzer for S3
- Block Public Access settings for this account
- Storage Lens
 - Dashboards
 - Storage Lens groups [New](#)
 - AWS Organizations settings
- Feature spotlight 5

► AWS Marketplace for S3

Account snapshot

Last updated: Nov 18, 2023 by Storage Lens. Metrics are generated every 24 hours. [Learn more](#)

Total storage	Object count	Average object size
22.1 MB	5	4.4 MB

You can enable advanced metrics in the ["default-account-dashboard"](#) configuration.

Buckets (5) Info

Buckets are containers for data stored in S3. [Learn more](#)

[Refresh](#) [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

Name	AWS Region	Access	Creation date
<input type="radio"/> myairline-script	US East (N. Virginia) us-east-1	Bucket and objects not public	November 13, 2023, 22:24:30 (UTC-05:00)
<input type="radio"/> myairline-query	US East (N. Virginia) us-east-1	Objects can be public	November 19, 2023, 19:46:34 (UTC-05:00)
<input type="radio"/> myairline-output-data	US East (N. Virginia) us-east-1	Bucket and objects not public	November 13, 2023, 22:23:48 (UTC-05:00)
<input type="radio"/> myairline-input-data	US East (N. Virginia) us-east-1	Bucket and objects not public	November 13, 2023, 22:23:25 (UTC-05:00)