

Aviator 2.0

Aviation Data Insights

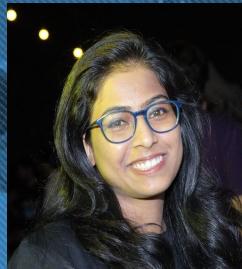




AGENDA

- Problem Statement
- Project Management
- Architecture
- RBAC
- Data Ingestion
- Snowflake Features
- Machine Learning
- Streamlit
- Security and Governance
- CI Implementation
- Dashboards
- Testing
- Documentation
- Future Scope

TEAM MEMBERS



Simran Jain
Mentor



Dikshant Jopat
Captain



Shreyas J
Sr. Software Engineer



Abhay Chilkawar
Sr. Software Engineer



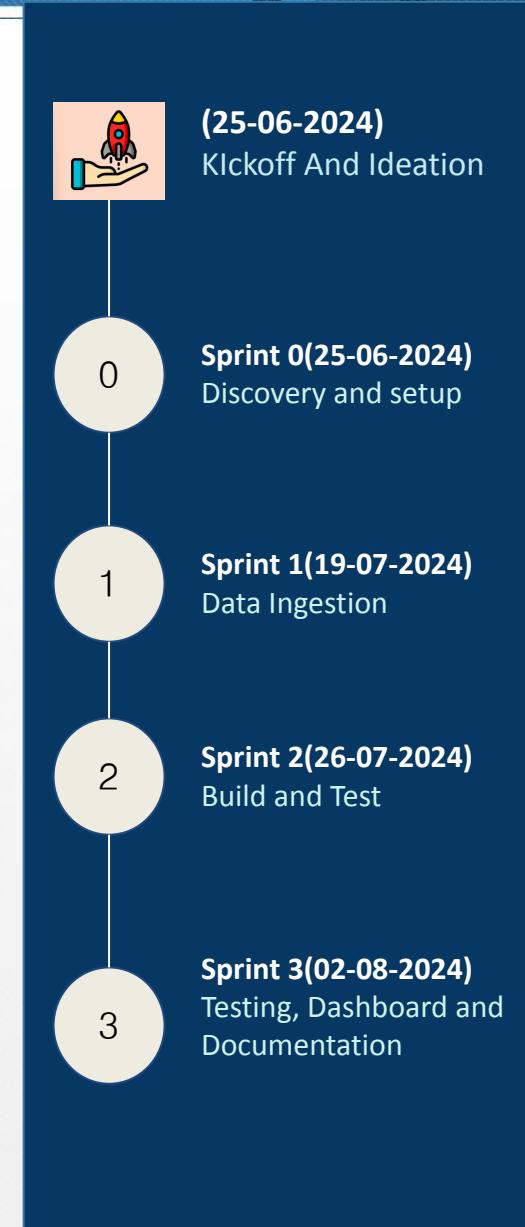
Neha Yerragudi
Software Engineer



Santoshi Sahu
Sr. Software Engineer

Problem Statement

- Flight delays are a major source of frustration for airline passengers, resulting in significant losses for airlines in terms of customer satisfaction and loyalty.
- Conducting flight customer satisfaction surveys is crucial for airlines to understand their passengers needs, preferences, and experiences.



PROPOSED SOLUTION



- Aviator is a company working with flight departure delay prediction using real-time weather data, along with analytics related to aviation industry.
- We also help airline companies to improve their services by analyzing and predicting the customer satisfaction based on customers feedback.
- Our model utilizes machine learning algorithms to forecast the likelihood of flight delays based on historical weather patterns, real-time weather conditions.
- Our flight delay prediction system helps airlines, airports, and travelers make informed decisions by providing accurate and timely predictions of flight delay probabilities, resulting in improved operational efficiency and enhanced customer experience.

PROJECT MANAGEMENT

1. **Project Plan**
2. **Daily Syncup**
3. **Slack Channel**
4. **Click up-Updates**

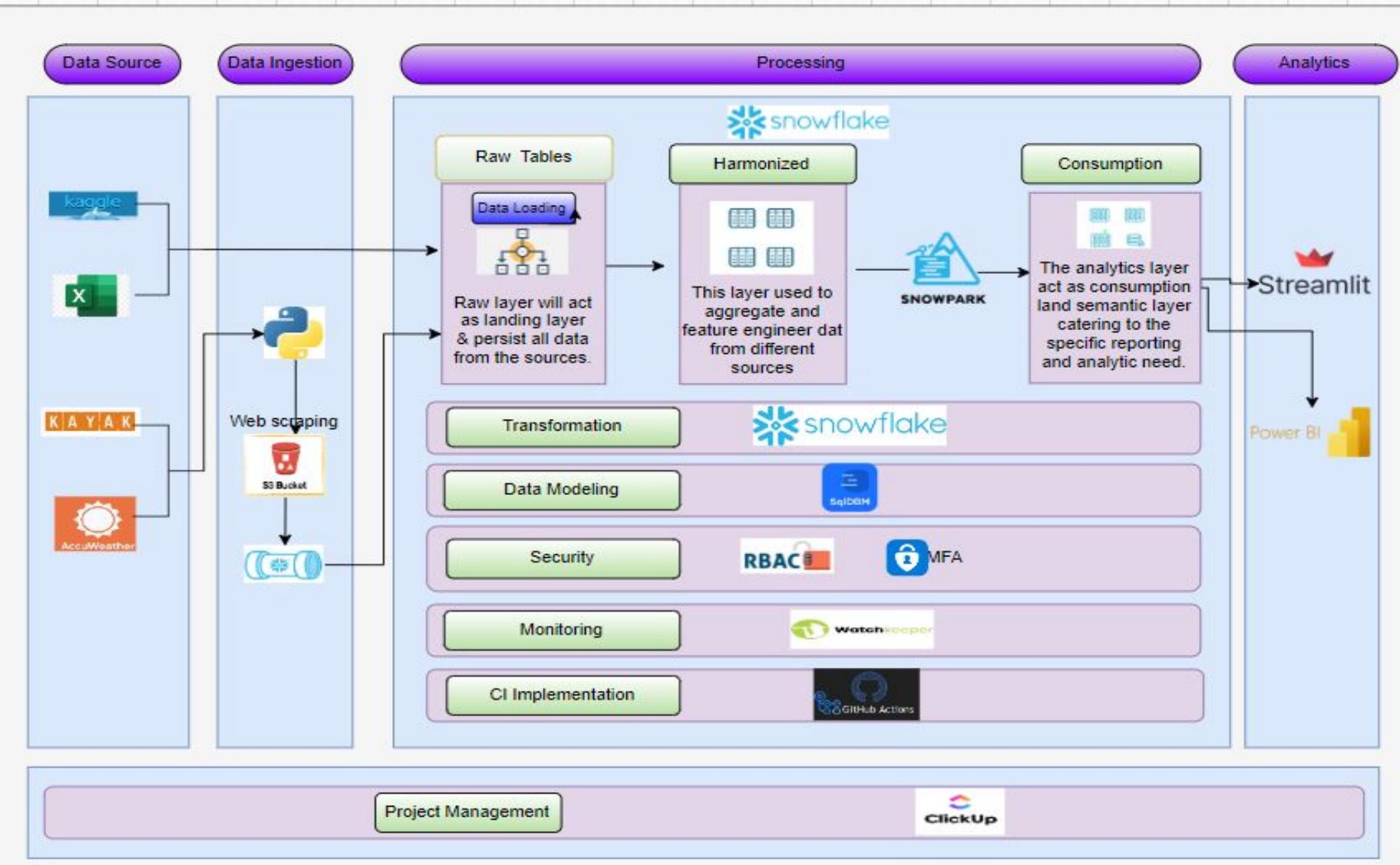
The screenshot shows the ClickUp interface for managing a project. On the left, there's a sidebar with navigation links like 'Docs', 'Dashboards', 'Clips', 'Timesheets', 'More', 'Favorites >', 'Spaces', 'Everything', 'Aviator' (selected), 'Discovery' (7 items), 'Development' (12 items, selected), 'Testing', 'Final', 'View all Spaces', 'Create Space', 'Invite', and a help icon.

The main area displays a Gantt chart for the month of August 2024. The chart shows several tasks under the 'Development' category, each with a color-coded timeline bar. The tasks include:

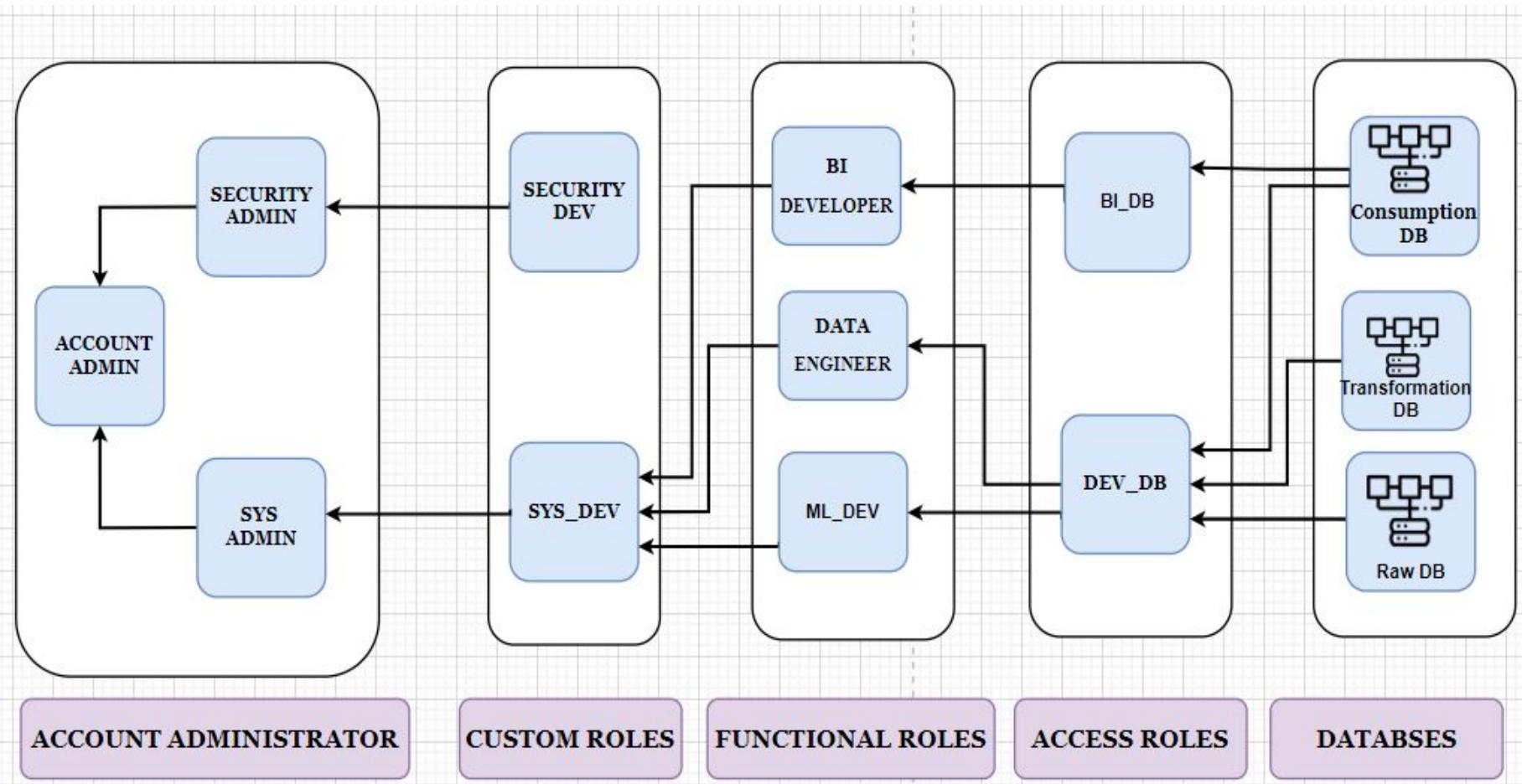
- Creation Role-Based Access Control (RBAC) in Snowflake (AC)
- Creation of Databases, schemas, and sch... (NY IC)
- Data loading into Raw Layer (SI)
- Data Transformation (SI)
- Development of ML Models (SI)
- CI/CD Integration with Github Actions (SI)
- Development of Data modeling using Sq... (SI)
- Creation of streamlit apps (SI)
- Building Dashboards in Powerbi (SI)

Each task has a detailed description and a small icon next to it. The 'Development' section in the sidebar is highlighted in purple, indicating it's the current workspace.

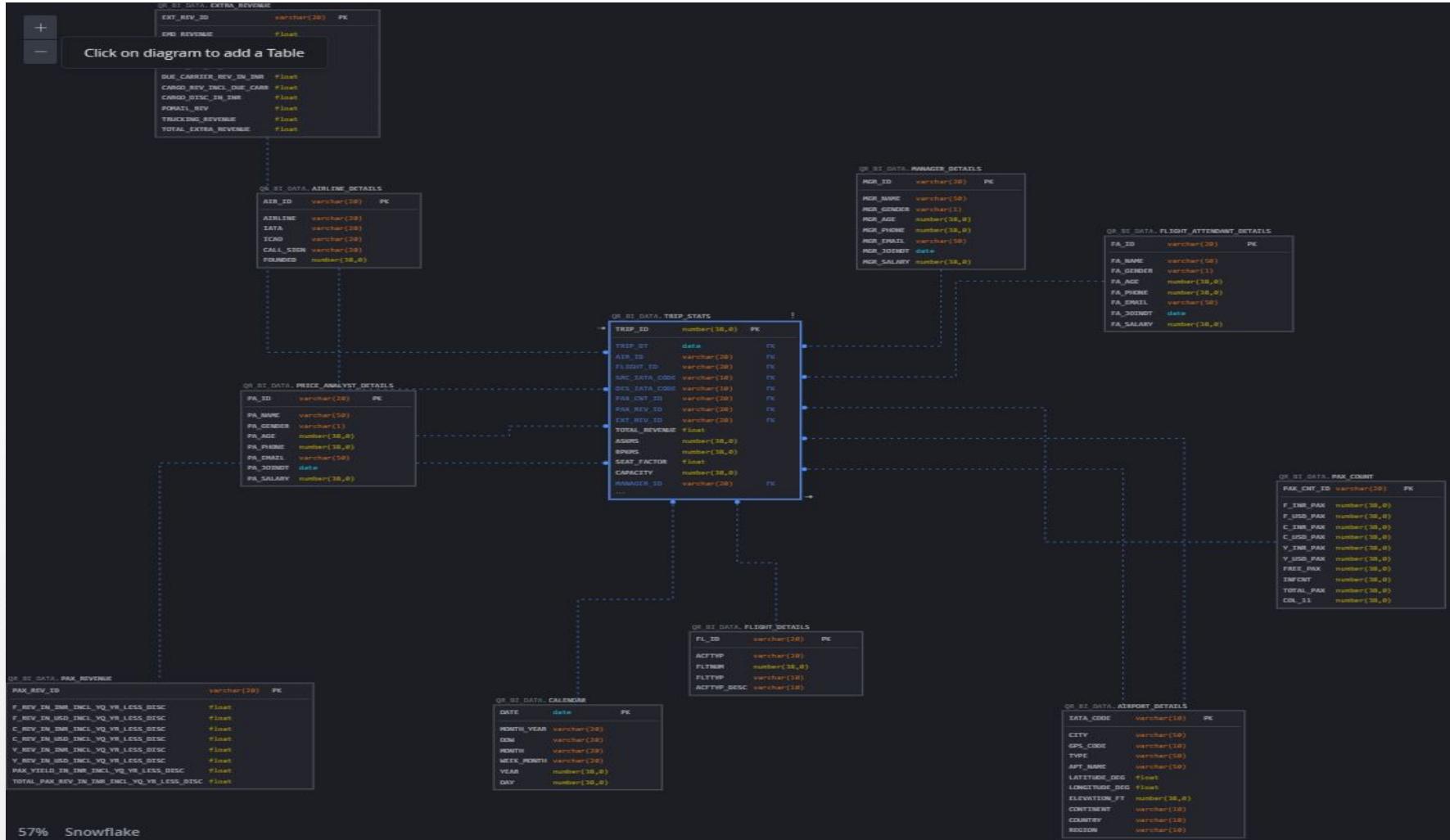
ARCHITECTURE



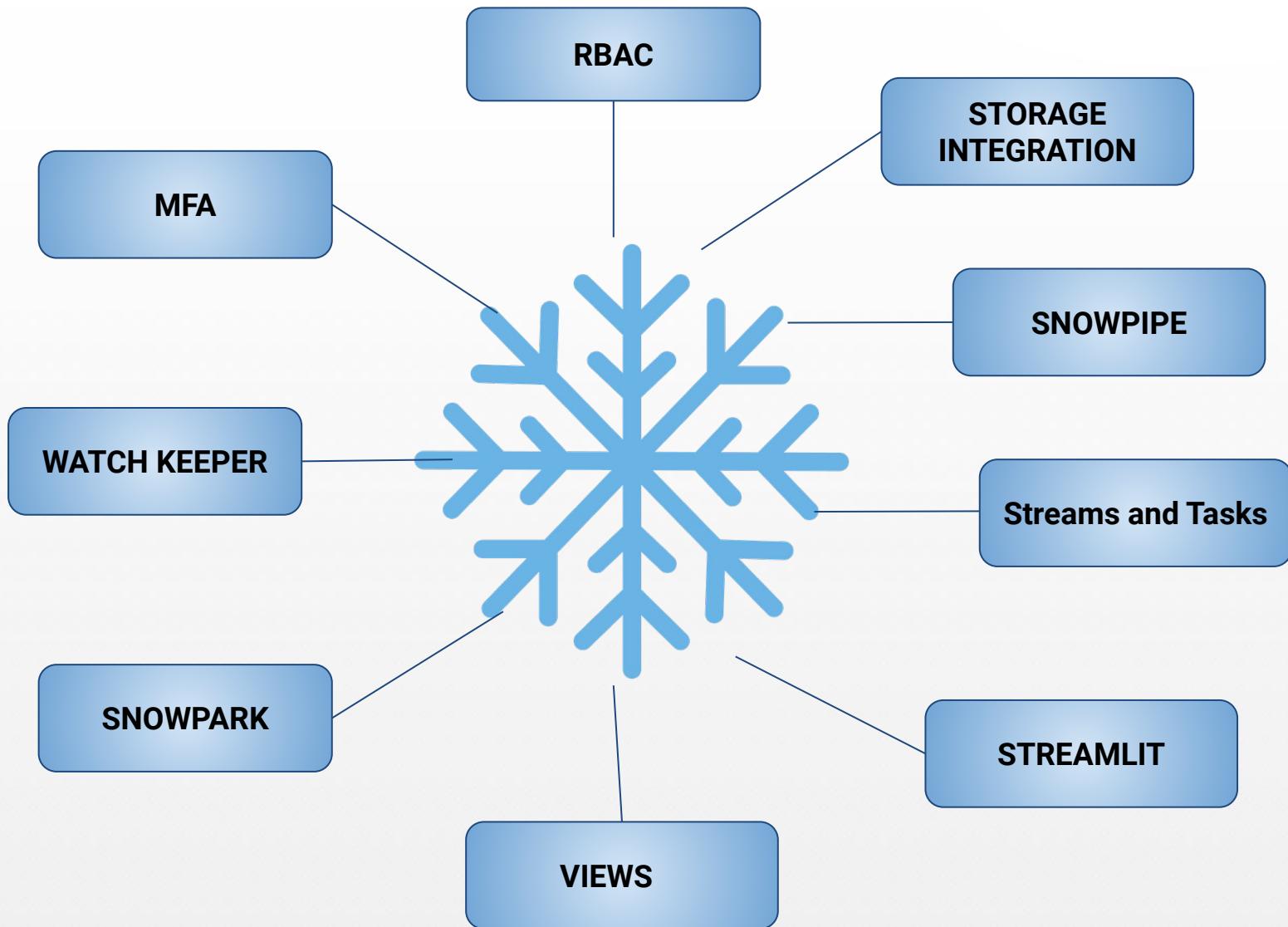
RBAC



DATA MODEL



FEATURES IMPLEMENTED ON SNOWFLAKE



Web scraping using Python



Python libraries used for web scraping

```
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
import pandas as pd
import time
from bs4 import BeautifulSoup as bs
from selenium.common.exceptions import NoSuchElementException
import re, json
from datetime import date, timedelta, datetime
import boto3
```

Web scraping using Python



To upload scraped data (json format) to **AWS S3 Bucket**

```
|  
aws_access_key_id = 'AKIA...' SUIGZVBVH4T'  
aws_secret_access_key = '7lpHK5tvDMNt..._CcXrD82CZ4AN6bJZv5TLMTYL'  
bucket_name = 'aviatorkipithon'  
# Create a session using your AWS credentials  
session = boto3.Session(  
    aws_access_key_id=aws_access_key_id,  
    aws_secret_access_key=aws_secret_access_key  
)  
  
# Create an S3 client  
s3_client = session.client('s3')  
  
# TODO implement  
optionss = webdriver.ChromeOptions()  
optionss.add_argument("start-maximized")  
optionss.add_experimental_option("excludeSwitches", ["enable-automation"])  
optionss.add_experimental_option('useAutomationExtension', False)  
#driver = webdriver.Chrome(ChromeDriverManager().install(), options=optionss)  
  
cService = webdriver.ChromeService(executable_path="C:\\\\Users\\\\ShreyasJ\\\\Desktop\\\\Kipithon\\\\chromedriver.exe")  
driver = webdriver.Chrome(service=cService,options=optionss)  
  
json_file_name = "Weather-" + str(str(int(date.today().day + day_num -1)) + '-' + str(date.today().month) + '-' + str(date.today().year))  
with open(json_file_name, "w") as json_file:  
    json.dump(listOfDict, json_file)  
  
s3_client.upload_file(json_file_name, bucket_name, 'weatherdata/{}'.format(json_file_name))
```

FILES IN S3 BUCKET



Amazon S3 > Buckets > [aviatorkipithon](#) > weatherdata/

weatherdata/

[Copy S3 URI](#)

Objects

Properties

Objects (7) [Info](#)



[Copy S3 URI](#)

[Copy URL](#)

[Download](#)

[Open](#)

[Delete](#)

[Actions](#) ▾

[Create folder](#)

[Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 > ⌂

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Weather-10-8-2024.json	json	August 6, 2024, 11:16:18 (UTC+05:30)	42.6 KB	Standard
<input type="checkbox"/>	Weather-11-8-2024.json	json	August 6, 2024, 11:16:19 (UTC+05:30)	43.0 KB	Standard
<input type="checkbox"/>	Weather-12-8-2024.json	json	August 6, 2024, 11:16:20 (UTC+05:30)	42.6 KB	Standard
<input type="checkbox"/>	Weather-6-8-2024.json	json	August 6, 2024, 11:16:14 (UTC+05:30)	43.0 KB	Standard
<input type="checkbox"/>	Weather-7-8-2024.json	json	August 6, 2024, 11:16:15 (UTC+05:30)	42.9 KB	Standard
<input type="checkbox"/>	Weather-8-8-2024.json	json	August 6, 2024, 11:16:16 (UTC+05:30)	42.8 KB	Standard
<input type="checkbox"/>	Weather-9-8-2024.json	json	August 6, 2024, 11:16:17 (UTC+05:30)	42.7 KB	Standard

Snowpipe and storage integration



```
CREATE OR REPLACE STORAGE INTEGRATION S3_AVIATOR
TYPE = EXTERNAL_STAGE
STORAGE_PROVIDER = S3
STORAGE_AWS_ROLE_ARN = 'arn:aws:iam::785781658896:role/Kipithon'
ENABLED = TRUE
STORAGE_ALLOWED_LOCATIONS = ('s3://aviatorkipithon/flightsdata/',
's3://aviatorkipithon/QR_DATA/','s3://aviatorkipithon/weatherdata');

CREATE TABLE RAW.DELAY_DATA.WEATHER_WEBSCRAPED (
RAW VARIANT
);

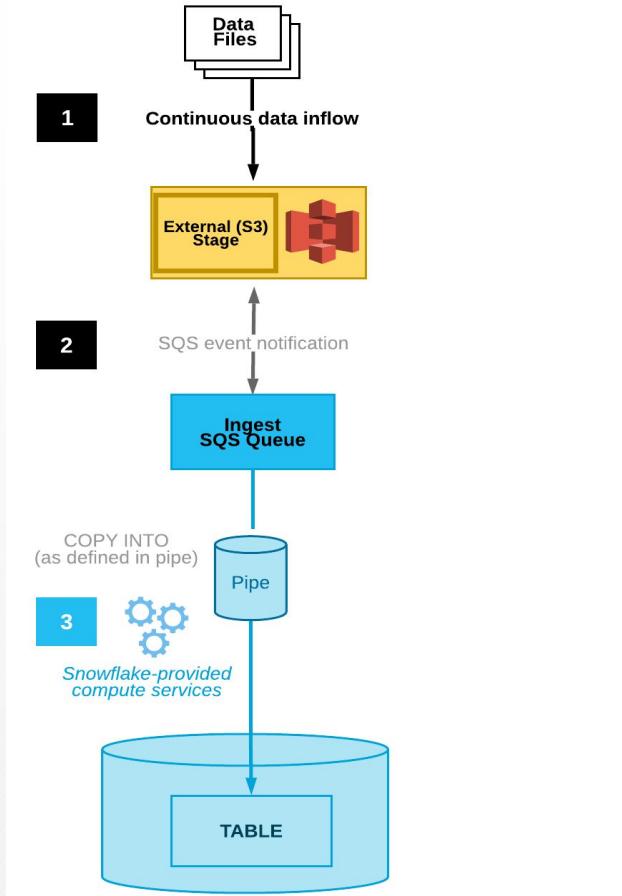
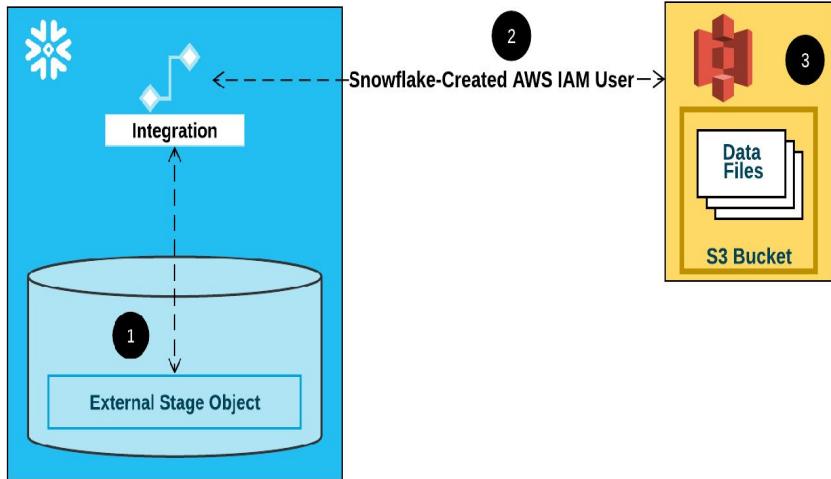
CREATE OR REPLACE STAGE RAW.DELAY_DATA.WEATHER_PIPE_STG
URL='s3://aviatorkipithon/weatherdata/'
STORAGE_INTEGRATION=S3_AVIATOR;

CREATE PIPE RAW.DELAY_DATA.weather_pipe
AUTO_INGEST = TRUE
AS
COPY INTO RAW.DELAY_DATA.WEATHER_WEBSCRAPED
FROM @RAW.DELAY_DATA.WEATHER_PIPE_STG
FILE_FORMAT = (TYPE = 'JSON');

show pipes;
select * from RAW.DELAY_DATA.WEATHER_WEBSCRAPED;

list @RAW.DELAY_DATA.WEATHER_PIPE_STG;
```

Snowpipe and storage integration



Storage Integration

Snowpipe

MODEL METHODOLOGY



1. DATASET COLLECTION



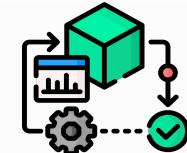
Aviation and Weather Data is collected from Kayak , Accuweather and Kaggle respectively

2. DATA PREPROCESSING



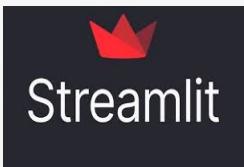
Data Cleaning, handling of missing data, outliers and Null records.

3. DATA TRAINING



Train ML Model to predict the probability of flight delay and Customer Satisfaction

6. DEPLOYMENT



Web Application for model for prediction of probability

5. MODEL EVALUATION



Determine trained model's effectiveness and quality using various metrics and approaches and select the top performing model

4. DATA TESTING



Test on design of a system for expected outputs for given inputs

ML Modeling



```
import snowflake.snowpark as snowpark
import pandas as pd
from pycaret.classification import *
import os,sys
from sklearn.model_selection import *

def main(session: snowpark.Session):
    merged_table = 'FEEDBACK_DATA.FEEDBACK_SCH.FEEDBACK_DATA_CLEAN'
    merged_df = session.table(merged_table).to_pandas()

    data= setup(merged_df,target='SATISFACTION', train_size = 0.8,fold=5,
               categorical_features=['GENDER','CUSTOMER_TYPE','TYPE_OF_TRAVEL','CLASS'],
               data_split_stratify=True,
               fix_imbalance=True,fix_imbalance_method = 'ADASYN', session_id=1,remove_outliers=True,
               ignore_features=['RW_NUMBER','ID'])
    top_models = compare_models(sort='AUC')

    train_metrics = pull()
    train_metrics_sp = session.create_dataframe(train_metrics)
    train_metrics_sp.write.mode("overwrite").save_as_table("FEEDBACK_MODEL_RESULTS")

    pred = predict_model(top_models)
    test_metrics = pull()
    test_metrics_sp = session.create_dataframe(test_metrics)
    test_metrics_sp.write.mode("overwrite").save_as_table("FEEDBACK_MODEL_RESULTS_TEST")

    merged_df_test_sp = session.create_dataframe(merged_df)
    merged_df_test_sp.write.mode("overwrite").save_as_table("FEEDBACK_TEST_DATA")
    import_dir = sys._xoptions.get("snowflake_import_directory")

    save_model(top_models, os.path.join(import_dir, '/tmp/FEEDBACK_model'))
    session.file.put(os.path.join(import_dir, '/tmp/FEEDBACK_model.pkl'),
                    "@MODEL_STAGE",
                    auto_compress=False,
                    overwrite=True
    )
```



KIPITHON | AVIATOR 2.0

Flight Departure Delay Prediction

[HOME](#) [EDA](#) [Features](#) [Model Performance](#) [Bulk Inferencing](#) [Single Inferencing](#)

Problem Statement

- Flight delays are a major source of frustration for an airline company, resulting in significant losses for airlines in terms of customer satisfaction and loyalty.
- One of the main factor affecting flight delay is weather. Understanding the affect of weather conditions during future flight take off schedules and taking precautionary measures are critical in ensuring smooth operations and high customer satisfaction.

Impact of solving the problem to the Business

- Improved Operational Efficiency: By predicting flight delays, airlines can proactively adjust their operations to minimize the impact of delays. This includes rerouting flights, re-assigning crew and aircraft, and adjusting ground staff schedules.
- Enhanced Customer Experience: Airlines can provide passengers with more accurate and timely information about flight delays, allowing them to plan accordingly. This can lead to increased customer satisfaction and loyalty.
- Reduced Costs: Flight delays can result in significant costs, including fuel, crew overtime, and passenger compensation. By predicting delays, airlines can take steps to mitigate these costs.



KIPITHON || AVIATOR

Airline Customer Satisfaction Analysis

[HOME](#) [EDA](#) [Features](#) [Model Performance](#) [Bulk Inferencing](#) [Single Inferencing](#)

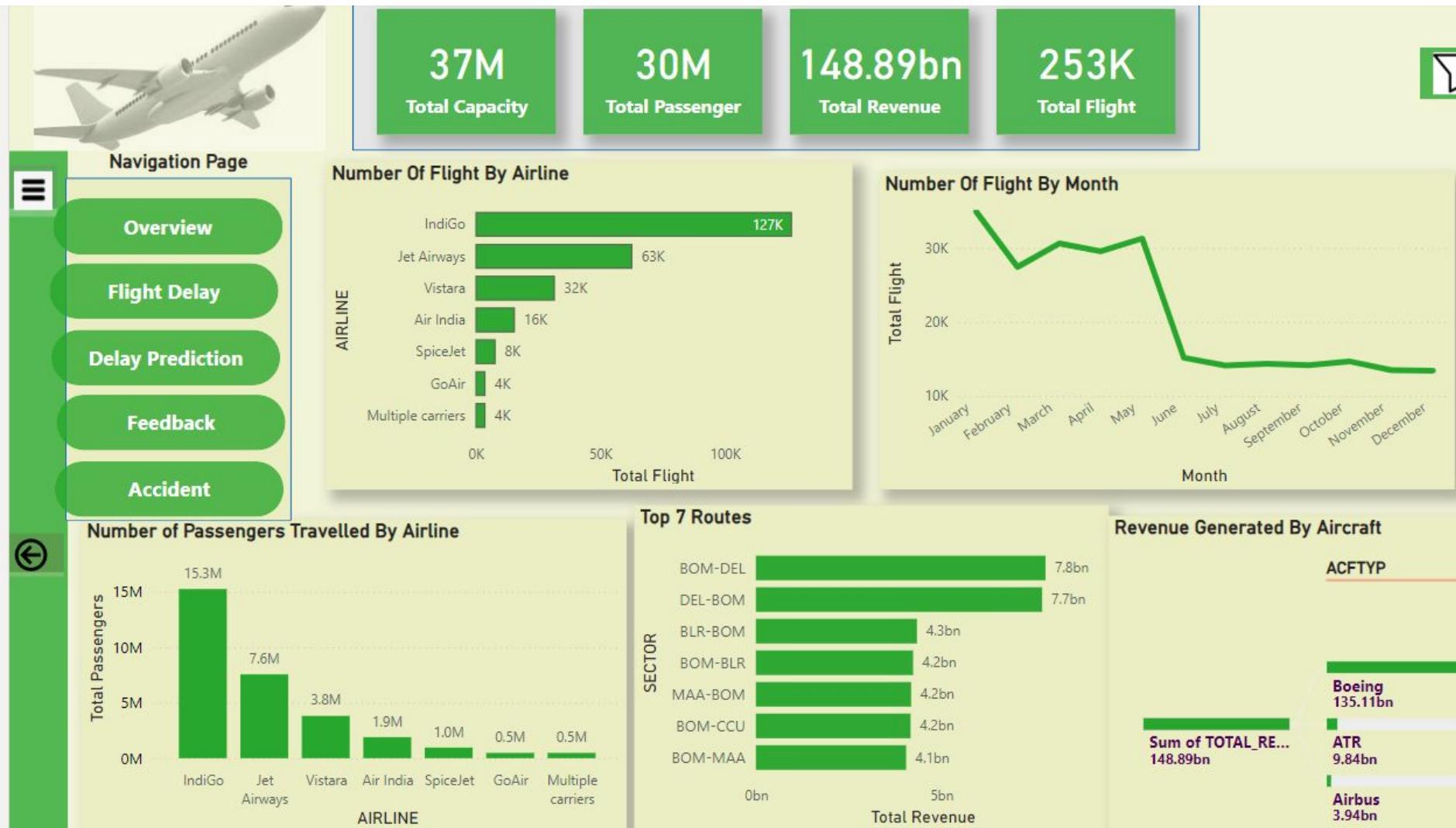
Problem Statement

- The airline industry is highly competitive, and customer satisfaction plays a crucial role in differentiating one airline from another.
- With the rise of social media and online review platforms, customers are more vocal than ever about their travel experiences.
- Analyzing customer feedback and sentiment can help airlines identify areas of improvement, increase customer loyalty, and ultimately, drive business growth.

Impact of solving the problem to the Business

- Enhance Customer Experience: Identify key drivers of customer satisfaction and improve overall customer experience, leading to increased loyalty and retention.
- Increase Revenue: Improve customer satisfaction ratings, leading to increased revenue through repeat business and positive word-of-mouth.
- Gain Competitive Advantage: Differentiate themselves from competitors by providing exceptional

POWER BI



Flight Delay

**28.27K**

Total Delayed Fl...

21.09

Average Temper...

36.95

Average wind

89.68

Average Cloud C...

24.43

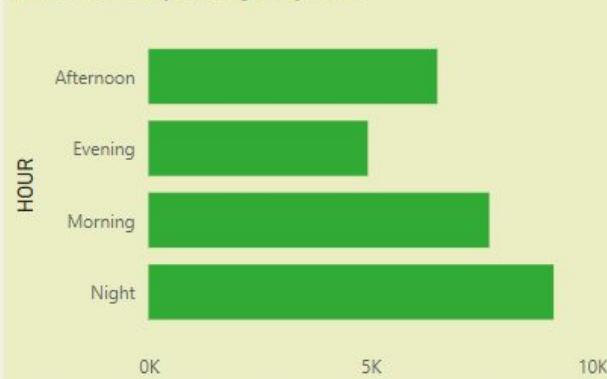
Average Precipi...



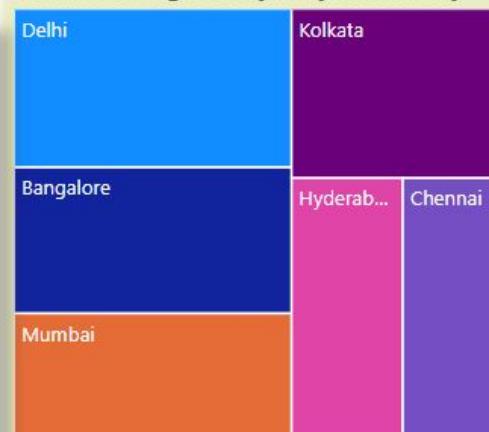
Navigation Page

[Overview](#)[Flight Delay](#)[Delay Prediction](#)[Feedback](#)[Accident](#)

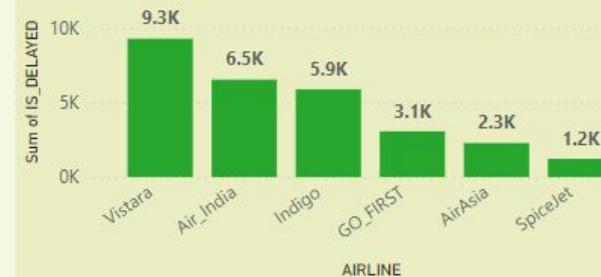
Count of Delayed Flight by Hour



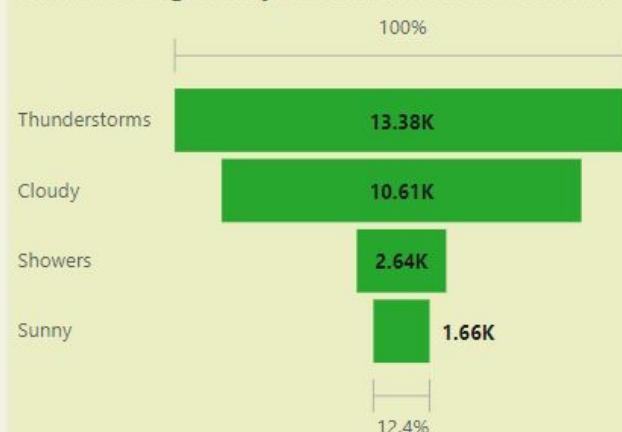
Number of Flight delayed by source City



Number Of Delayed Flight By Airline



Number of Flight delayed based on Weather condition



Delay Prediction



Report view



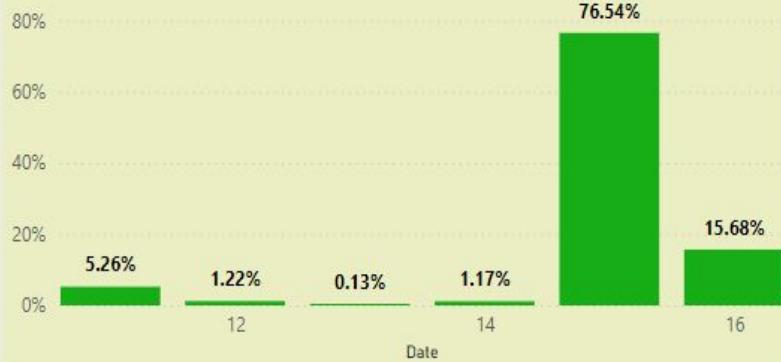
SOURCE_CITY

Bengaluru	Hyderabad	Mumbai
Chennai	Kolkata	

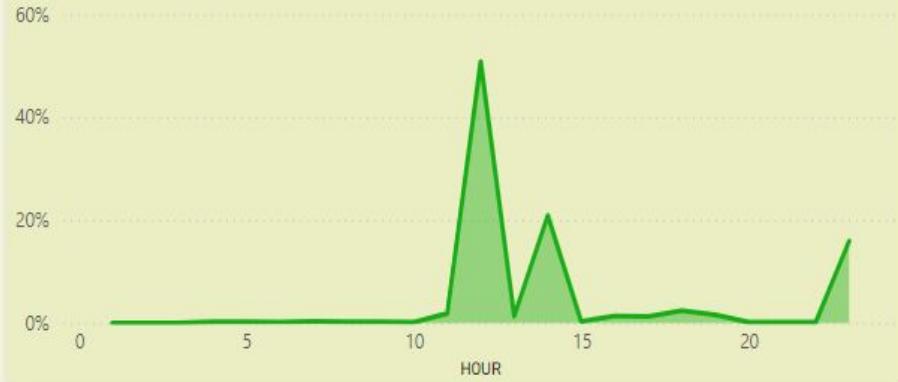
DESTINATION

Chennai	Hyderabad	Mumbai
Delhi	Kolkata	

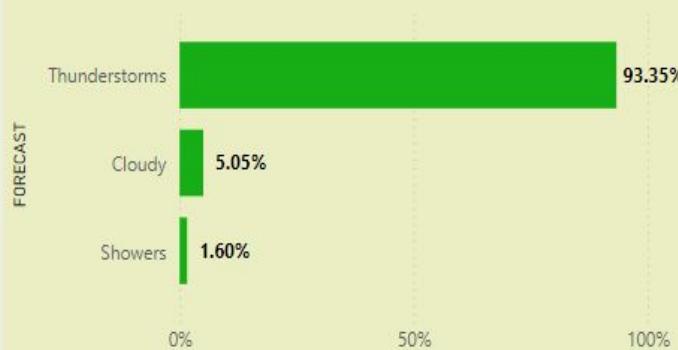
Delay Probability By Date



Delay Probability Over Hour



Forecast Of The Day



Weather Conditions Of The Day



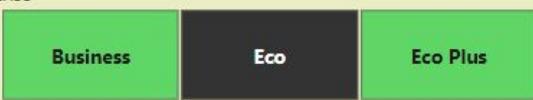
Parameter

- CLOUD_CO...
- PRECIPITAT...
- TEMPERAT...
- VISIBILITY...
- WIND_RAN...

Feedback



CLASS



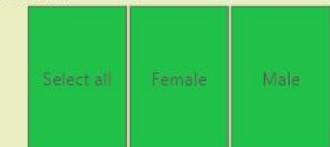
TYPE_OF_TRAVEL



CUSTOMER_TYPE

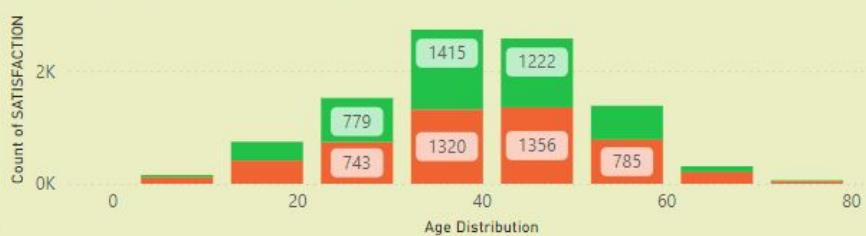


GENDER



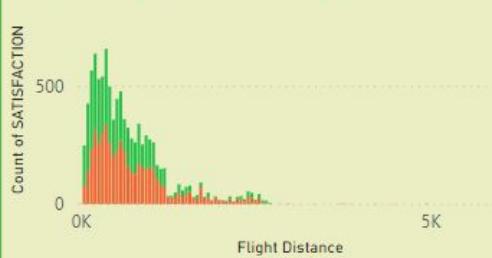
Satisfaction Count By Age Group

SATISFACTION ● neutral or dissatisfied ● satisfied

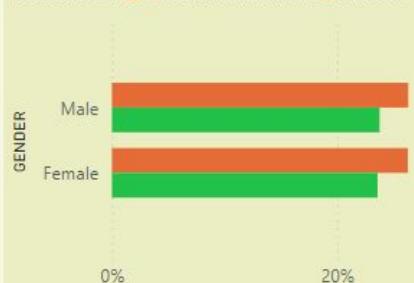


Satisfaction Count By Distance

SATISFACTION ● neutral or dissatisfied ● satisfied



SATISFACTION ● neutral or dissatisfied ● satisfied



Baggage handling



Food and drink



Leg Room Service



Checkin Service



Gate Location



Inflight wifi service



Inflight Entertainment



Cleanliness



Departure Arrival Tim...



Ease of online...



Accident



33K

Total Fatal Injur...

22K

Total Minor Inju...

17K

Total Serious Inj...

72K

Total Injuries

Navigation Page

Overview

Flight Schedule

Flight Delay

Feedback

Accident

Engine.Type

● Reciprocating

● Turbo Fan

● Turbo Shaft

● Turbo Prop

● Turbo Jet

● LR

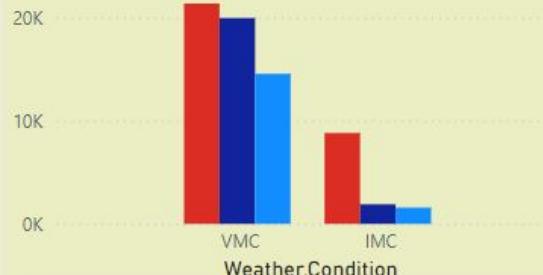
● Electric

● Hybrid Roc...

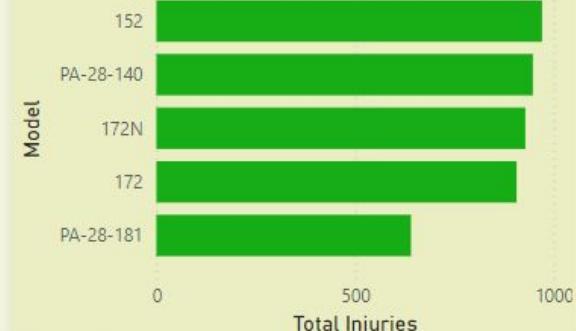
● Geared Tur...

Injuries Based On Weather Condition

● Sum of Total.Fatal.Inj... ● Sum of Total.... ● Sum of Total....



Top 5 Models Caused More Injuries



Total Injuries By Year



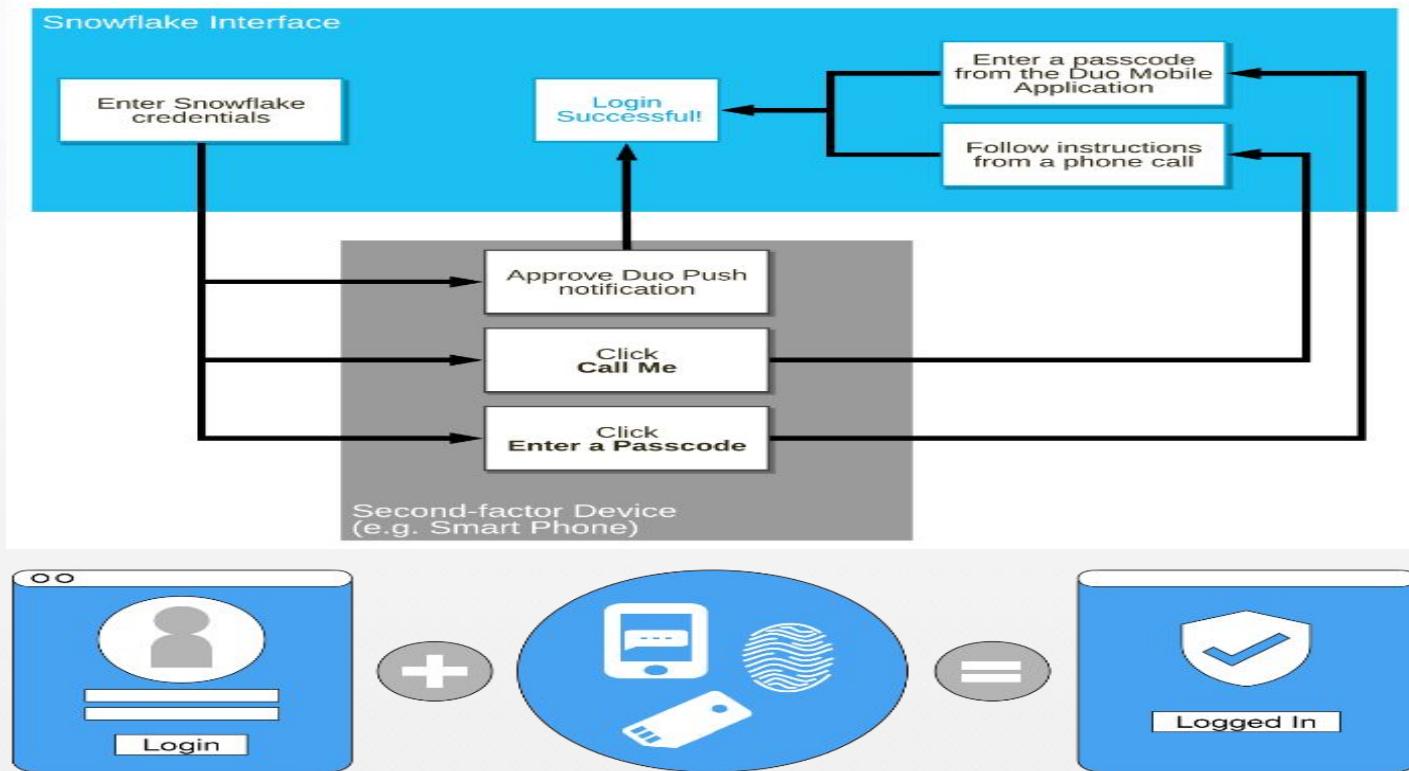
Total Injuries By Purpose Of flight



Security and Governance



- In Snowflake, Multi-Factor Authentication (MFA) is an additional security layer that requires users to provide a second form of verification, beyond just a username and password, to access their Snowflake account.
- This adds an extra layer of protection against unauthorized access, phishing, and other types of attacks.



Multi Factor Authentication



- Account: KJ27319

 **Snowflake**

Select an account to sign into

KJ27319

WK92745

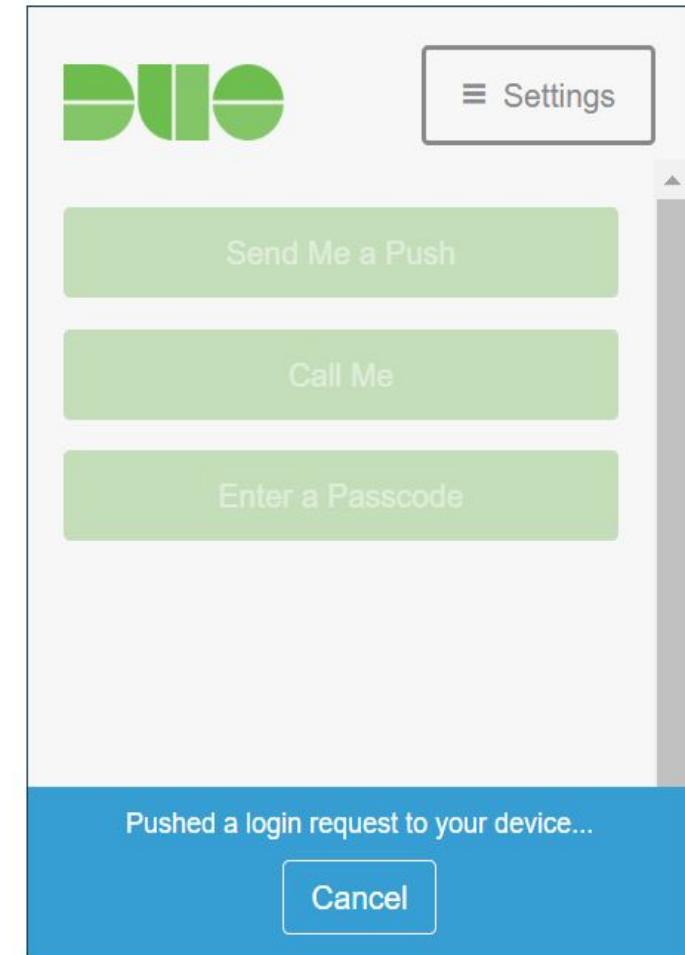
mab93037

Other accounts ▾

[Sign into a different account](#)

or

Sign up



CI IMPLEMENTATION



- CI/CD is a modern software development practice in which incremental code changes are made frequently and reliably.
- Implemented our **Continuous Integration** using Github Actions.

Screenshot of a GitHub repository page for "Kipithon-Aviator".

The repository is private and has 1 branch and 0 tags. The main branch is selected. The repository was last updated 4 minutes ago by user "yerragudiNeha" with commit "consumption updates". Other commits include "Update and rename main.yml to UPDATED.yml" and "consumption updates". Initial commits for "LICENSE" and "README.md" were also made.

The repository has 21 commits in total. It includes a "Readme" file and an "Apache-2.0 license". There is 1 watcher, 0 forks, and 0 stars. The repository has not published any releases.

Code | Issues | Pull requests | Actions | Projects | Security | Insights | Settings

Kipithon-Aviator (Private)

Unwatch 1 | Fork 0 | Star 0

main | 1 Branch | 0 Tags

Go to file | Add file | Code

yerragudiNeha consumption updates · abbf4e6 · 4 minutes ago · 21 Commits

.github/workflows · Update and rename main.yml to UPDATED.yml · 1 hour ago

SNOWFLAKE_OBJECTS · consumption updates · 4 minutes ago

LICENSE · Initial commit · 1 hour ago

README.md · Initial commit · 1 hour ago

README · Apache-2.0 license

About

No description, website, or topics provided.

Readme | Apache-2.0 license | Activity | 0 stars | 1 watching | 0 forks

Releases

No releases published | Create a new release

Packages

Kipithon-Aviator

GITHUB ACTIONS AND JOBS



Secrets Variables

Repository secrets

New repository secret

Name	Last updated	Action
SF_ACCOUNT	1 hour ago	
SF_DATABASE	1 hour ago	
SF_PASSWORD	1 hour ago	
SF_ROLE	1 hour ago	
SF_SCHEMA	1 hour ago	
SF_URL	1 hour ago	
SF_USERNAME	1 hour ago	
SF_WAREHOUSE	1 hour ago	

✓ schema commit
snowflake-devops #9: Commit 8c4b6c0 pushed by yerragudiNeha main

✓ initail commit
snowflake-devops #8: Commit 32830cb pushed by yerragudiNeha main

✓ commit
snowflake-devops #4: Commit 04148f4 pushed by yerragudiNeha main

✓ db commit
snowflake-devops #3: Commit 8e4a58e pushed by yerragudiNeha main

✓ Merge branch 'main' of https://github.com/yerragu...
snowflake-devops #2: Commit 00b4ddb pushed by yerragudiNeha main

✓ snowflake-devops
snowflake-devops #1: Manually run by yerragudiNeha main

Files

main

Go to file

.github

SNOWFLAKE_OBJECTS

- V1.1.1_initial_objects.sql
- V1.1.2_raw_layer.sql
- V1.1.4_Transform_layer_objects...
- V1.1.5_consumption_layer.sql
- v1.1.3_Transformlayer.sql

Kipithon-Aviator / SNOWFLAKE_OBJECTS /



yerragudiNeha consumption updates

Name	Last commit message
..	
V1.1.1_initial_objects.sql	Tables creation
V1.1.2_raw_layer.sql	tables creation
V1.1.4_Transform_layer_objects.sql	Transforming data
V1.1.5_consumption_layer.sql	consumption updates

TESTING



- **Testing** can be stated as the process of verifying and validating whether our product is bug-free, meets the technical requirements as guided by its design and development, and meets the user requirements effectively.

... > 04 Delivery kipithon > Testing and Validation ▾

Type ▾ People ▾ Modified ▾

Name ↑	Owner	Last mo... ▾
Dashboard Testing	me	11:02 AM
Integration Testing	S Santoshi Sahu	8:07 PM
Test execution-Data Loading	me	Aug 6, 2024
Test scenarios	me	Aug 6, 2024

Snowflake Scenarios testing



A	B	C	D	E	
1	Requirement	Test Scenario ID	Test Scenario Name	Test Steps	Test Cases
2	Snowflake Scenarios	TS_01	Data Load Test	Step 1 Step 2 Step 3	Verify that data is loaded accurately and completely into Snowflake from various sources. Test different files, such as CSV (Static & Dynamic data used) to ensure proper loading. Test scenarios with large data volumes to evaluate performance and scalability.
3					
4					
5		TS_02	Security Test	Step 1 Step 2	Validate user access controls and permissions to ensure data confidentiality and integrity. Test various authentication methods, such as username/password, multi-factor authentication, or integration with identity providers.
6					
7		TS_03	Data transformation Test	Step 1 Step 2 Step 3	Test data transformation logic implemented in Snowflake, such as ETL (Extract, Transform, Load) processes. Verify data integration between Snowflake and external systems, such as data warehouses or third-party tools. Validate data consistency and accuracy after transformation and integration processes.
8					
9					
10		TS_04	Error Handling	Step 1	Test the recovery process after system failures, such as node failures or network interruptions.
11		TS_05	Scalability Testing	Step 1 Step 2 Step 3	Simulate concurrent user access to evaluate system performance and response time. Test scenarios with high data ingestion rates or concurrent data modifications to assess system limits. Measure the impact of resource allocation, such as warehouses or virtual warehouses, on concurrent operations.
12					
13					
14		TS_06	Data Retention	Step 1 Step 2	Test data retention policies and verify that data is archived or purged according to the defined rules. Ensure data availability and integrity.
15					

Models & PowerBi testing



	A	B	C	D	E
16	Dynamic Delay prediction	TS_01	Delay prediction and Accuracy	Step 1	Test the accuracy of delay prediction based on weather condition such as visibility, wind etc.
17				Step 2	Verify that the delay prediction algorithm is functioning correctly and producing accurate results for different flights.
18		TS_02	user Interface	Step 1	Test the user interface and user experience related to dynamic delay
19				Step 2	Verify that the displayed delays are updated in real-time as per the dynamic delay algorithm.
20		TS_03	Delay Change Timing and Frequency	Step 1	Test how quickly and frequently Delay change in response to dynamic factors
21				Step 2	Verify that delay changes are triggered in a timely manner.
22		TS_04	Competitive Analysis	Step 1	Perform competitive analysis by monitoring and comparing Delays with competitors in the industry.
23					
24					
25	Airline customer satisfaction analysis	TS_01	customer satisfaction analysis	Step 1	Test different customer rating based on different services such as wifi, food, cleanliness, age group .
26					
27	Power BI Scenariois	TS_01	Delay Accuracy and Completeness	Step 1	Verify that the data extracted from the snowflake is accurately represented in the Power BI dashboard.
28				Step 2	Cross-check the delay prediction in the dashboard with the actual prediction .
29		TS_02	Real-Time Updates	Step 1	Verify that the dashboard reflects the accurate delay prediction for different airlines and routes.
30				Step 2	Test scenarios with rapid delay changes to ensure the dashboard updates in a timely manner.
31		TS_04	Visualizations and Charts	Step 1	Verify that the visualizations, such as line charts, bar charts, or tables, accurately represent the data.
32				Step 2	Test different combinations of filters and slicers to ensure that the visualizations dynamically update based on user input.
33				Step 3	Check the interactivity and responsiveness of the visualizations when interacting with them.

DOCUMENTATION



My Drive > Kipithon/Aviator (Diksha...)

Type People Modified

Name	Owner	Last modified	File size	⋮
01 Initiation	me	Aug 5, 2024	—	⋮
02 Discovery	me	Aug 5, 2024	—	⋮
03 Planning	me	Aug 5, 2024	—	⋮
04 Delivery kipithon	me	Aug 5, 2024	—	⋮
05 Project Management	me	Aug 5, 2024	—	⋮
06 References	me	10:17 AM	—	⋮

My Dri... > Kipithon/Aviator (Diks... > O4 Delivery kip

Type People Modified

Name	Owner
Data Engineer	me
Source Code	me
Testing and Validation	me
Visualisation	S Santoshi Sahu
Architechture.png	me
Aviation Data Insights.pptx	me

FUTURE SCOPE



- Extending Flight delay Prediction to predict accurate delay time in minutes.
- Find out top n parameters from services provided by airlines that hinders the growth in revenue.
- Including other factors such as airport traffic, flight conditions, crew availability along with weather conditions to predict delay.
- Integration of customer satisfaction with recommendation systems to recommend certain discounts for customers based on their satisfaction.



THANK YOU

