



Hotel Bookings Analysis Status Report

Group – 18

Jeyadev L

Gayathiri R

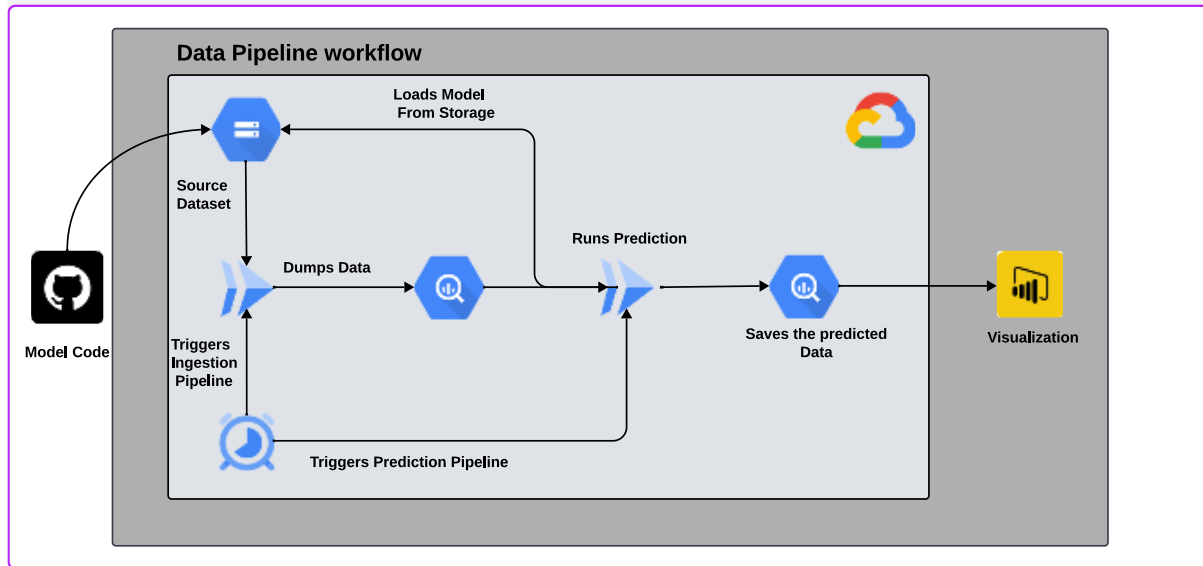
Bratati R

Devasree R

Sharone V

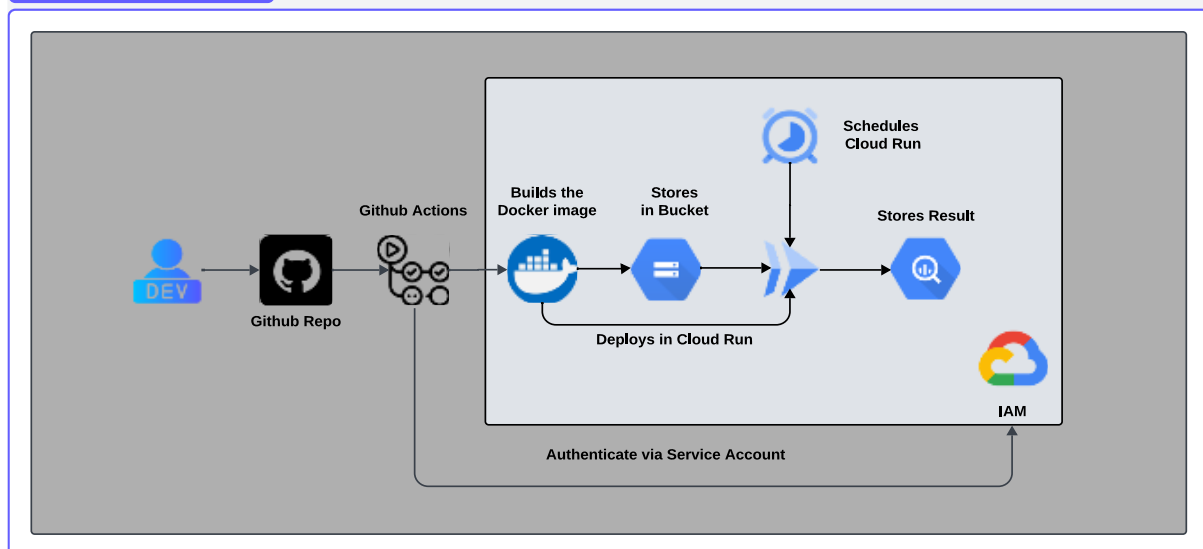
Data Pipeline Architecture Diagram

Pipeline Workflow



Deployment Architecture Diagram

Deployment Workflow Diagram



Agile Board

Group 18 Space / Capstone Project / Hotel Bookings Analysis and Optimization

Ask AI Share Automations

Hotel Bookings Analysis and Optimization

Board List Gantt Table View

COMPLETE 19

+

Add Task

Name	Assignee	Start date	Due date	Priority	Status	Comments	Sprint points
📄 Creating Document for standardization ...	👤	📅	Oct 18	Normal	COMPLETE	🗨️ 1	...
📄 Handle Missing Values & Standardiz...	👤	Oct 15	Oct 17	Urgent	COMPLETE	🗨️ 1	...
📄 Load and Explore the Dataset	👤	Oct 15	Oct 17	Urgent	COMPLETE	🗨️ 1	...
📄 Detect and Correct Outliers & Gener...	👤	Oct 15	Oct 18	High	COMPLETE	🗨️ 1	...
📄 EDA	👤	Oct 18	Oct 19	📅	COMPLETE	🗨️ 0	...
📄 Creating Synthetic data	👤	Oct 18	📅	📅	COMPLETE	🗨️ 0	...
📄 Creating a Data stream	👤	Oct 18	📅	High	COMPLETE	🗨️ 0	...
📄 Data Exploration	👤	Oct 23	Oct 24	Urgent	COMPLETE	🗨️ 0	...
📄 Data Prep For ML Model	👤	Oct 22	Oct 26	📅	COMPLETE	🗨️ 0	...
📄 Phase - 1	👤	Oct 15	Oct 15	📅	COMPLETE	🗨️ 0	...
📄 Decide on Forecasting Method	👤	Oct 24	Oct 30	High	COMPLETE	🗨️ 0	...
📄 Feature Engineering	👤	Oct 22	Oct 26	📅	COMPLETE	🗨️ 0	...
📄 Model Implementation	👤	Oct 22	Oct 26	📅	COMPLETE	🗨️ 0	...
📄 Model Selection	👤	Oct 26	📅	📅	COMPLETE	🗨️ 0	...
📄 Phase - 2	👤	Oct 21	Nov 4	Normal	COMPLETE	🗨️ 0	...
📄 Setting Up Database	👤	Oct 22	Nov 12	📅	COMPLETE	🗨️ 0	...
📄 Explore Deployment Cloud Options - Ar...	👤	Oct 21	Nov 12	📅	COMPLETE	🗨️ 0	...
📄 Model Documentation	👤	📅	📅	📅	COMPLETE	🗨️ 0	...
📄 Deploy to Cloud	👤	Oct 28	📅	📅	COMPLETE	🗨️ 0	...

+ Add Task

Save

Group 18 Space / Capstone Project / Hotel Bookings Analysis and Optimization

Ask AI Share Automations

Hotel Bookings Analysis and Optimization

Board List Gantt Table View

COMPLETE 19

+

Add Task

IN PROGRESS 4

+

Add Task

Name	Assignee	Start date	Due date	Priority	Status	Comments	Sprint points
📄 Phase - 3	👤	Nov 4	📅	Low	IN PROGRESS	🗨️ 0	...
📄 Data Pipeline - Arch - Document	👤	Oct 24	Nov 30	📅	IN PROGRESS	🗨️ 0	...
📄 Phase - 4	👤	4 days ago	📅	📅	IN PROGRESS	🗨️ 0	...
📄 Data Quality - Check & Alerts	👤	📅	📅	📅	IN PROGRESS	🗨️ 0	...

+ Add Task

Save

BACKLOG 4

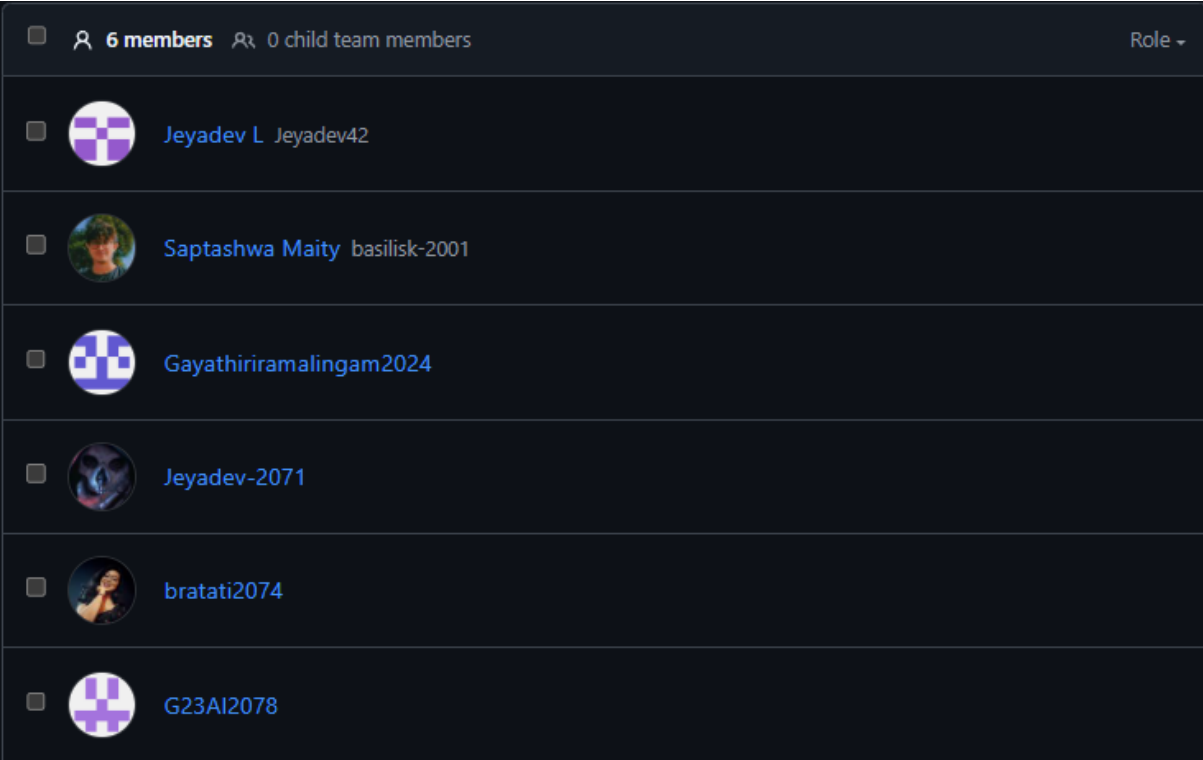
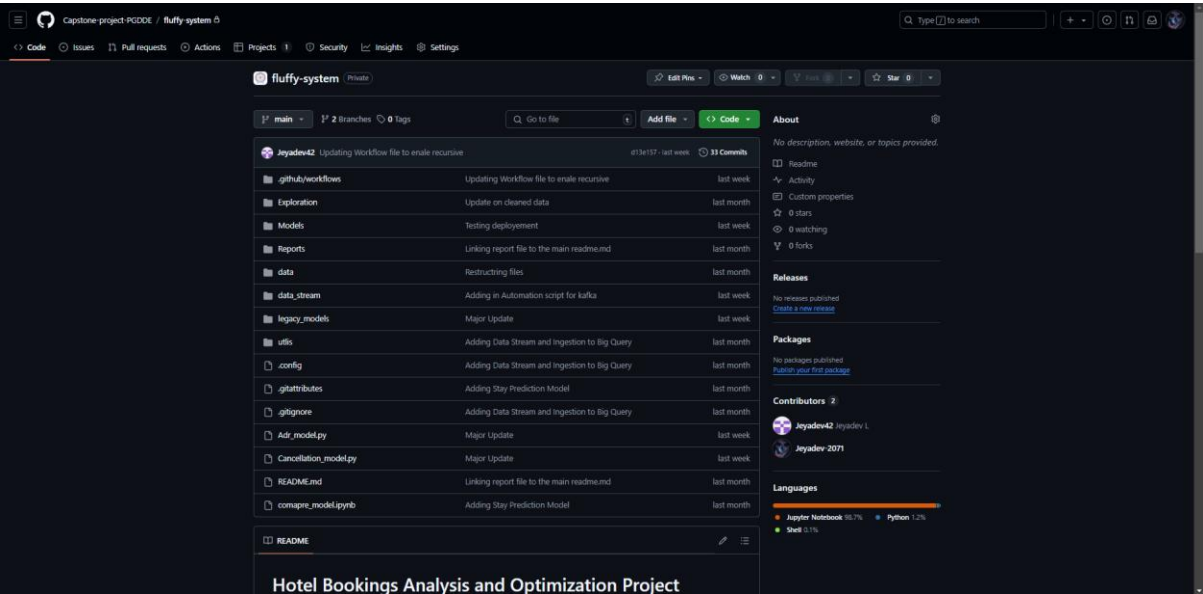
+

Add Task

+ Add Task

Save

GitHub Setup



Cloud Setup

Database :-

Google Cloud

Capstone-project-group-18

Bucket

Search

Explorer

Q Search BigQuery resources

SHOW STARRED ONLY

capstone-project-group-18

Queries

Notebooks

Data canvases

Data preparations

Workflows

External connections

CAPSTONE_PROJECT

HOTEL_BOOKING

PREDICTION_RESULTS

DDL

CREATE SCHEMA 'itij-capstone-project-group-18' CAPSTONE_PROJECT ;

CREATE TABLE 'CAPSTONE_PROJECT' . 'HOTEL_BOOKING' (

guid STRING,

hotel STRING,

is_canceled INTEGER,

lead_time INTEGER,

arrival_date_year INTEGER,

arrival_date_month STRING,

arrival_date_week_number INTEGER,

arrival_date_day_of_month INTEGER,

stays_in_weekend_nights INTEGER,

stays_in_week_nights INTEGER,

adults INTEGER,

children FLOAT64,

babies INTEGER,

meal STRING,

country STRING,

market_segment STRING,

distribution_channel STRING,

is_repeated_guest INTEGER,

previous_bookings_not_canceled INTEGER,

reserved_room_type STRING,

assigned_room_type STRING,

booking_changes INTEGER,

deposit_type STRING,

agent STRING,

company STRING,

days_in_waiting_list INTEGER,

customer_type STRING,

adr FLOAT64,

required_car_parking_spaces INTEGER,

total_of_special_requests INTEGER,

reservation_status STRING,

reservation_status_date STRING,

insert_timestamp TIMESTAMP

);

CREATE OR REPLACE TABLE 'itij-capstone-project-group-18' . CAPSTONE_PROJECT . PREDICTION_RESULTS (

guid STRING,

lead_time INTEGER,

hotel STRING,

market_segment STRING,

previous_cancellations INTEGER,

booking_changes INTEGER,

total_of_special_requests INTEGER,

arrival_date_month STRING,

SUMMARY

ACTIVITY

Job history

Job history

Model Storage Cloud

Secure your account and protect your data by turning on multi-factor authentication. Add an extra layer of security today.

LEARN MORE

Bucket details

GO TO PATH

REFRESH

LEARN

model-storage-group-18

Location

Storage class

Public access

Protection

Hierarchical namespace

Location

Storage class

Public access

Protection

Hierarchical namespace

Objects

Configuration

Permissions

Protection

Lifecycle

Observability

Inventory reports

Operations

Folder browser

model-storage-group-18

Models

model-storage-group-18

Models

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

Name

Size

Type

Created

Storage class

Last modified

Public access

Version history

Encryption

gradient_boosting_best_model.pkl

19.1 MB

application/octet-stream

Nov 14, 2024, 7:12:51 PM

Standard

Nov 14, 2024, 7:12:51 PM

Not public

Google-managed

gradient_boosting_regressor_best_model.pkl

186 KB

application/octet-stream

Nov 14, 2024, 7:12:48 PM

Standard

Nov 14, 2024, 7:12:48 PM

Not public

Google-managed

Docker Image

Google CloudCapstone-project-group-18Art

Images for gcr.ioDELETEEDIT REPOSITORYSETUP INSTRUCTIONSREFRESH

gcr.io > itj-capstone-project-group-18

Repository Details

Format: DockerType: StandardSHOW MORE

Filter

Enter property name or value

Name	Connection	Created	Updated
data-pipeline-service	—	6 hours ago	2 hours ago

Cloud Run Setup

Google CloudCapstone-project-group-18Cloud

Cloud RunService detailsEDIT & DEPLOY NEW REVISIONSET UP CONTINUOUS DEPLOYMENTLEARN

data-pipeline-service

Region: asia-south1URL: https://data-pipeline-service-880519683680-asia-south1.run.appService min instances: 0

METRICS

SLOS

LOGS

REVISIONS

TRIGGERS

NETWORKING

SECURITY

YAML

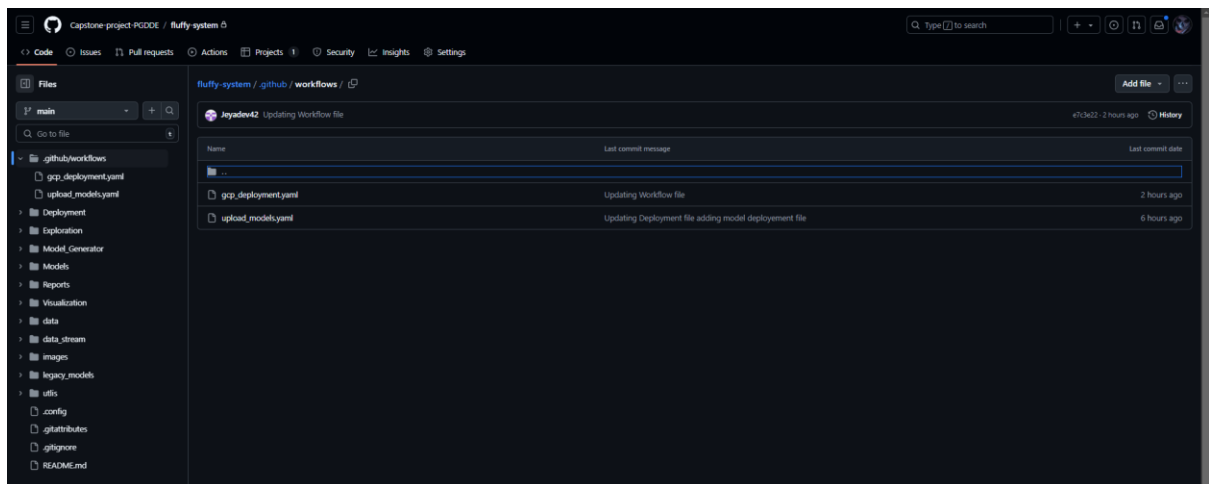
EDIT

```
1 apiVersion: serving.knative.dev/v1
2 kind: Service
3 metadata:
4   name: data-pipeline-service
5   namespace: '880519683680'
6   selfLink: /apis/serving.knative.dev/v1/namespaces/880519683680/services/data-pipeline-service
7   uid: 87e393ad-f5d1-473f-822b-3ac6d6a2a8d8
8   resourceVersion: AAY8KdF418
9   generation: 0
10  creationTimestamp: '2024-11-28T05:29:03.651966Z'
11  labels:
12    run.googleapis.com/satisfiesPps: 'true'
13    cloud.googleapis.com/location: asia-south1
14  annotations:
15    serving.knative.dev/creator: 880519683680-compute@developer.gserviceaccount.com
16    serving.knative.dev/lastModifier: 880519683680-compute@developer.gserviceaccount.com
17    run.googleapis.com/client-name: gcloud
18    run.googleapis.com/client-version: 502.0.0
19    run.googleapis.com/operation-id: 40a36037-beta-413d-943d-196ff4c366a
20    run.googleapis.com/ingress: all
21    run.googleapis.com/ingress-status: all
22    run.googleapis.com/url: '["https://data-pipeline-service-880519683680-asia-south1.run.app","https://data-pipeline-service-7nafrjvrs-el-a.run.app"]'
23  spec:
24    template:
25      metadata:
26        labels:
27          client.knative.dev/name: appt4pdxvc
28          run.googleapis.com/startupProbeType: Default
29      annotations:
30        autoscaling.knative.dev/maxScale: '100'
31        run.googleapis.com/client-name: gcloud
32        run.googleapis.com/client-version: 502.0.0
33        run.googleapis.com/startup-cpu-boost: 'true'
34      spec:
35        containerConcurrency: 80
36        timeoutSeconds: 120
37        serviceAccountName: 880519683680-compute@developer.gserviceaccount.com
38        containers:
39          - image: gcr.io/itj-capstone-project-group-18/data-pipeline-service:latest
40            ports:
41              - name: http
```

Scheduler

	Cloud Scheduler	Jobs	CREATE JOB	REFRESH	FORCE RUN	EDIT	COPY	PAUSE	RESUME	DELETE	LEARN
SCHEDULER JOBS APP ENGINE CRON JOBS											
Filter Filter jobs											
<input type="checkbox"/>	Name ↑	Status of last execution	Region	State	Description	Frequency	Target	Last run	Next run	Last updated	Actions
<input type="checkbox"/>	capstone-cancellation-prediction-job	Success	asia-south1	Enabled		0 */5 * * * * (Asia/Calcutta)	URL: https://data-pipeline-service-880519683680.asia-south1.run.app/predict	Nov 28, 2024, 12:00:00 PM	Nov 28, 2024, 6:00:00 PM	Nov 28, 2024, 11:33:53 AM	⋮
<input type="checkbox"/>	capstone-ingestion-job	Success	asia-south1	Enabled		0 * * * * (Asia/Calcutta)	URL: https://data-pipeline-service-880519683680.asia-south1.run.app/ingest	Nov 28, 2024, 5:00:00 PM	Nov 28, 2024, 6:00:00 PM	Nov 28, 2024, 11:34:11 AM	⋮

GitHub workflow set up



```
name: Upload Models to GCS

on:
  push:
    paths:
      - 'Models/**' # Trigger the workflow only when files in the Models folder change

env:
  PROJECT_ID: ${ secrets.GCP_PROJECT_ID } # Google Cloud Project ID
  BUCKET_NAME: ${ secrets.GCS_BUCKET_NAME } # Google Cloud Storage bucket name
  MODELS_PATH: Models/ # Path to the models folder

jobs:
  upload-models:
    runs-on: ubuntu-latest

    steps:
      # Step 1: Checkout the code
      - name: Checkout code
        uses: actions/checkout@v3

      # Step 2: Authenticate with Google Cloud
      - name: Authenticate with Google Cloud
        uses: google-github-actions/auth@v1
        with:
          credentials_json: ${ secrets.GCP_SERVICE_ACCOUNT_KEY }

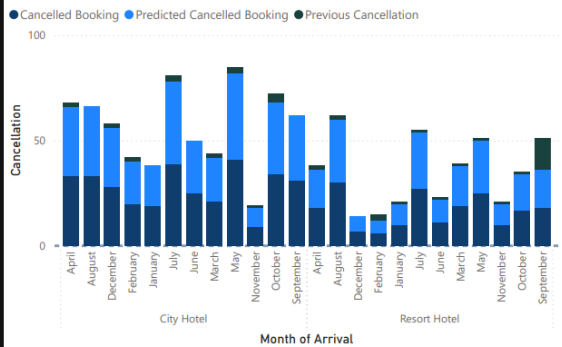
      # Step 3: Configure Google Cloud SDK
      - name: Configure Google Cloud SDK
        run: |
          gcloud config set project $PROJECT_ID

      # Step 4: Upload updated or new files to GCS
      - name: Upload Models to GCS
        run: |
          for file in $(git diff --name-only --diff-filter=AM HEAD~1 | grep "^${MODELS_PATH}"); do
            echo "Uploading $file to GCS bucket $BUCKET_NAME"
            gsutil cp "$file" "gs://$BUCKET_NAME/$file"
          done
```

```
1  name: Build and Deploy to Cloud Run
2
3  on:
4    push:
5      branches:
6        - main
7      paths:
8        - 'Models/**'
9        - 'Deployment'
10
11  env:
12    PROJECT_ID: ${ secrets.GCP_PROJECT_ID }
13    SERVICE_NAME: data-pipeline-service
14    REGION: asia-south1
15
16  jobs:
17    build-and-deploy:
18      runs-on: ubuntu-latest
19
20      steps:
21        # Step 1: Checkout repository
22        - name: Checkout code
23          uses: actions/checkout@v3
24
25        # Step 2: Authenticate with Google Cloud
26        - name: Authenticate with Google Cloud
27          uses: google-github-actions/auth@v1
28          with:
29            credentials_json: ${ secrets.GCP_SERVICE_ACCOUNT_KEY }
30
31        # Step 3: Configure Docker for GCP
32        - name: Configure Docker for GCP
33          run: gcloud auth configure-docker
34
35        # Step 4: Debug file paths (optional)
36        - name: Debug Files
37          run: ls -R
38
39        # Step 5: Build Docker Image
40        - name: Build Docker Image
41          run: |
42            docker build -t gcr.io/$PROJECT_ID/$SERVICE_NAME:latest -f Deployment/Dockerfile .
43
44        # Step 6: Push Docker Image to GCR
45        - name: Push Docker Image
46          run: docker push gcr.io/$PROJECT_ID/$SERVICE_NAME:latest
47
48        # Step 7: Deploy to Cloud Run
49        - name: Deploy to Cloud Run
50          run: |
51            gcloud run deploy $SERVICE_NAME \
52              --image gcr.io/$PROJECT_ID/$SERVICE_NAME:latest \
53              --region $REGION \
54              --platform managed \
55              --allow-unauthenticated
```


PowerBI Report

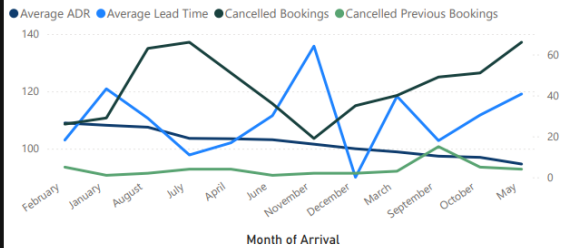
Cancellation Chart



Average Revenue Chart



Cancellation VS Lead Time VS ADR



Average Revenue By Market Segment

