MITIGATING FLIGHT DELAYS THROUGH WEATHER ANALYTICS

Team 1

EEET2574 – Big Data for Engineering

Assignment 3: Big Data Project

Team Members

Phan Nhat Minh
s3978598

- Project Manager
- Data Scientist

Tran Manh Cuong
s3974735

- Data Engineer
- Data Scientist

Nguyen Vinh Gia Bao
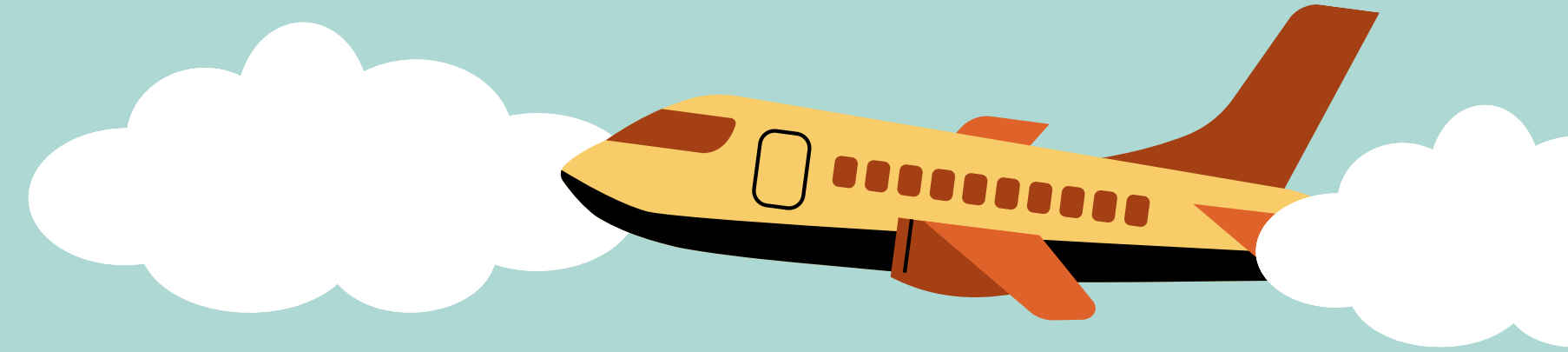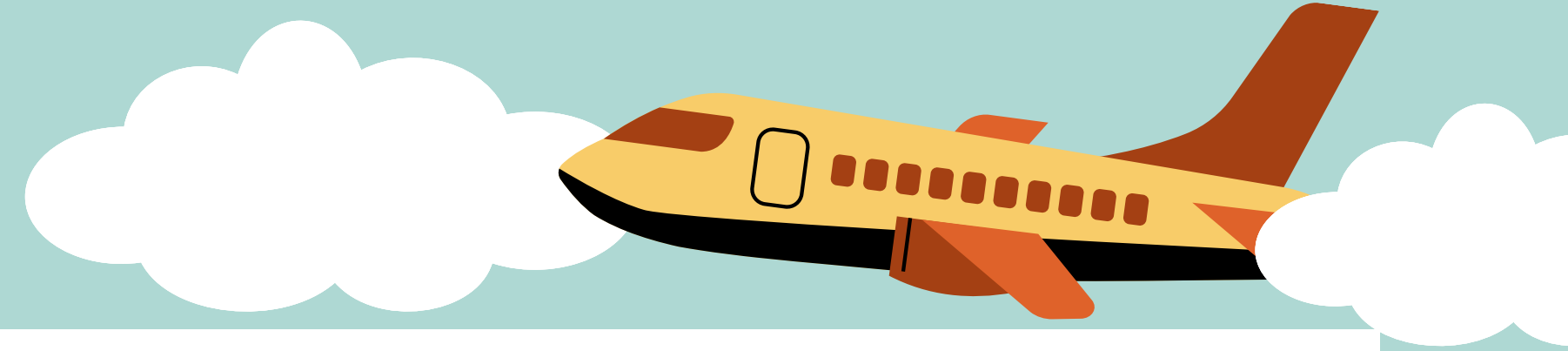s3986287

- Data Engineer
- Data Scientist

Introduction

Table 0-1: Direct cost of air transportation delay in 2007

| Cost Component	Cost ($ billions)
Costs to Airlines	8.3
Costs to Passengers	16.7
Costs from Lost Demand	3.9
Total Direct Cost	28.9
Impact on GDP	4.0
Total Cost	32.9

Introduction

Flight delays may **weaken the performance of the transportation system**, and will bring adverse impact on the planning of an airport.

Airlines are liable to **sanctions**, **fines** and **additional costs** caused by flight delays.

Introduction
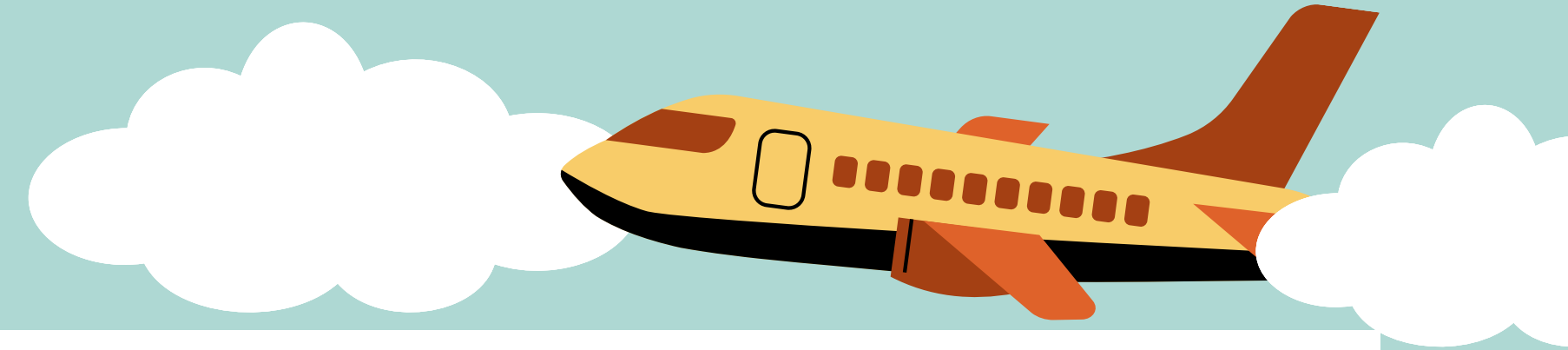
**Impact of Flight Delays on Passenger:**
Disrupted itineraries
Time and productivity loss
Emotional and physical stress
Financial implications
Reduced customer satisfaction
Physical discomfort
Missed events and special occasions
Health and well-being
Inconvenience for families and groups
Loss of trust and loyalty
Safety concerns
Negative ripple effects
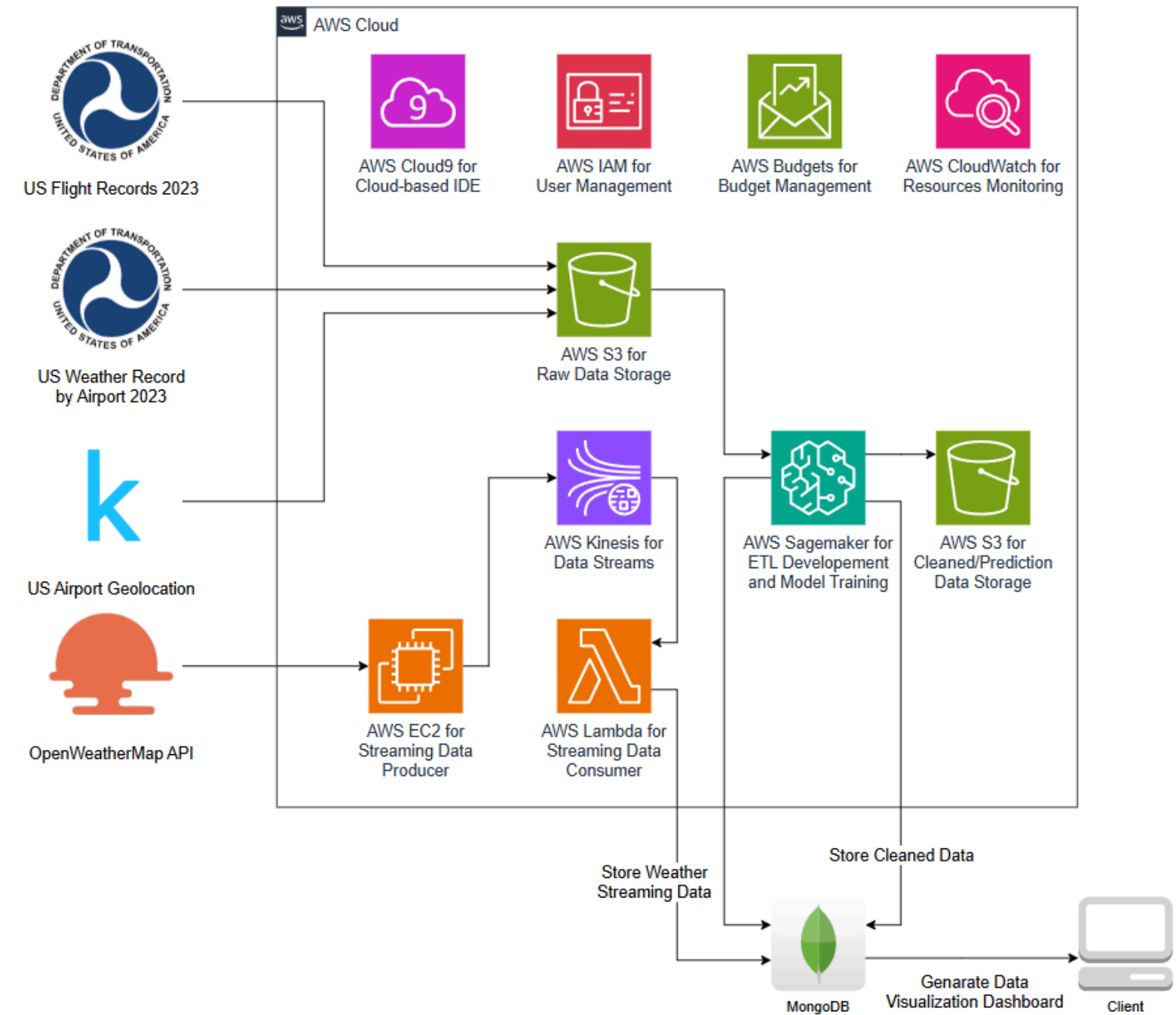
ESTIMATED COST

Mitigating Flight Delays through Weather Analytics and Big Data Edit ✎

Export ▼ Share

Estimate summary Info

Upfront cost	Monthly cost	Total 12 months cost
0.00 USD	131.48 USD	**1,577.76 USD** Includes upfront cost

Getting Started with AWS

Get started for free   Contact Sales

Mitigating Flight Delays through Weather Analytics and Big Data

Duplicate   Delete   Move to   Create group   Add support   **Add service**

Q Find resources

< 1 >

	Service Name ▲		Status ▽	Upfront cost ▽	Monthly cost ▽	Description ▽	Region ▽	Config Summary ▽
☐	Amazon EC2	✎	-	0.00 USD	4.23 USD	AWS EC2 for Streaming Data Producer	US East (N. Virginia)	Tenancy (Shared Instances), Operating system (Linux), Workload (Consistent, Number of instanc...
☐	Amazon Kinesis Data Streams	✎	-	0.00 USD	29.22 USD	AWS Kinesis for Data Streams	US East (N. Virginia)	Number of days for data retention (1 days), Number of records (48 per hour), Number of Consu...
☐	Amazon SageMaker	✎	-	0.00 USD	97.92 USD	AWS Sagemaker for ETL Development and Model Training	US East (N. Virginia)	Instance name (ml.c5.large), Number of data scientist(s) (2), Number of Studio Notebook instanc...
☐	Amazon Simple Storage Service (S3) ✎		-	0.00 USD	0.05 USD	AWS S3 for Raw Data Storage	US East (N. Virginia)	S3 Standard storage (2 GB per month), PUT, COPY, POST, LIST requests to S3 Standard (100), GE...
☐	Amazon Simple Storage Service (S3) ✎		-	0.00 USD	0.05 USD	AWS S3 for Cleaned/Prediction Data Storage	US East (N. Virginia)	S3 Standard storage (2 GB per month), PUT, COPY, POST, LIST requests to S3 Standard (100), GE...
☐	AWS Lambda	✎	-	0.00 USD	0.01 USD	AWS Lambda for Streaming Data Consumer	US East (N. Virginia)	Architecture (x86), Amount of ephemeral storage allocated (512 MB), Architecture (x86), Invoke ...

PRELIMINARY RESULT

Model Name	Test Accuracy	F1 Score	Precision	Recall
GBTClassifier	0.883478	0.876245	0.888864	0.883478
DecisionTreeClassifier	0.881242	0.873269	0.887953	0.881242
RandomForestClassifier	0.873855	0.863551	0.884591	0.873855
LogisticRegression	0.873214	0.862819	0.883926	0.873214

=> GBTClassifier was chosen as the model to predict whether the departure flight is delayed or not

PRELIMINARY RESULT

6 visualization dashboards:

- Overview Dashboard of the Airport

- Weather Impact on Delay Dashboard

- Geographical Insights Dashboard

- Aircraft Analysis Dashboard

- Time Series Dashboard

- Performance Benchmark

Demonstration