**Experiential Learning Approach for Deloitte**

# **Case Study 2 Report**

# Perform Uber Trip Cost Analytics: Functional Design

## Introduction

Uber is a technology company whose mission is to reimagine the way the world moves for the better. The technology at Uber helps to develop and maintain multisided platforms that match consumers looking for rides and independent providers of ride services, as well as with other forms of transportation, including public transit, bikes, and scooters.

Uber also connect consumers and restaurants, grocers, and other merchants so they can buy and sell meals, groceries, and other items, then these are matched with independent delivery service providers. Plus, Uber connects shippers and carriers in the freight industry.

The technology at Uber helps people connect and move in over 70 countries and 10,000 cities around the world.

## Problem statement

1. To capture, manage and store the real time data which is generated continuously since the company is providing the services to millions of people.
2. To Generate logs Continuously using the AWS Services.
3. To design an efficient algorithm using machine learning model which will predict the fare to be charged.
4. To visualize the Analyzed & Cleaned data to gain insights.

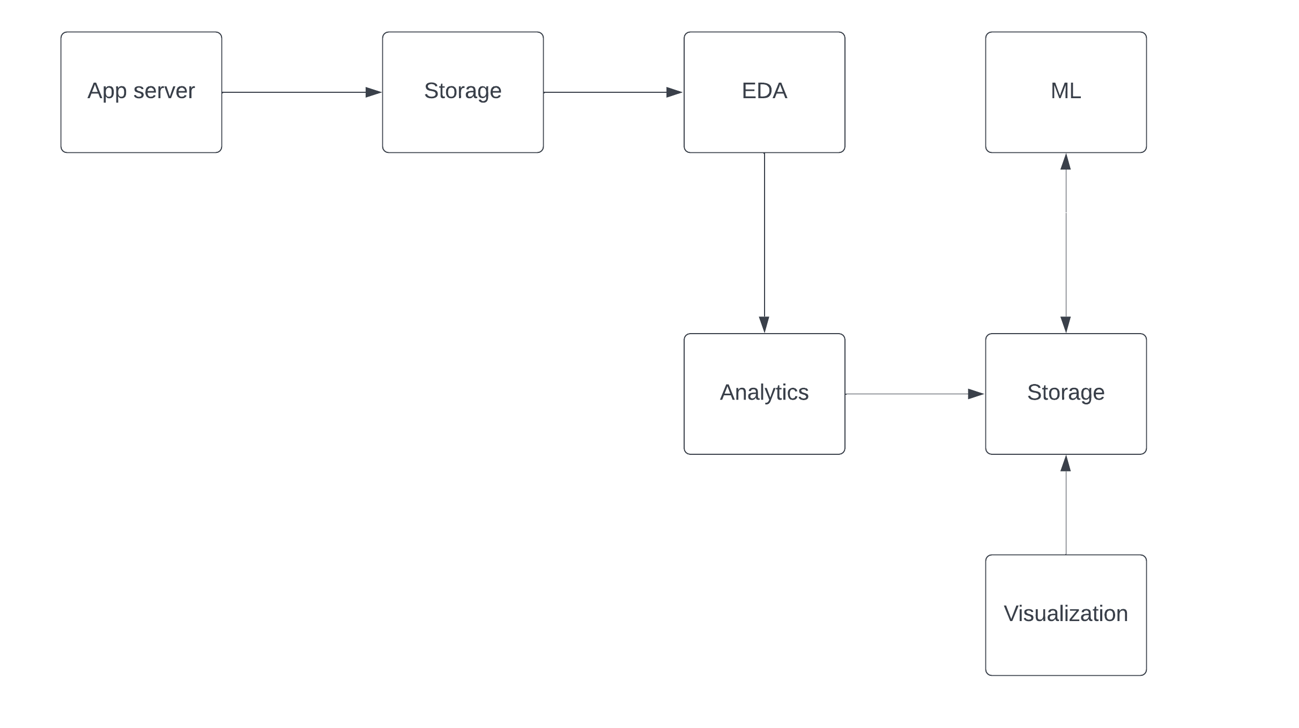
## Description of data provided

Since the data is generated in real time, the columns of the data to be received are as follows:

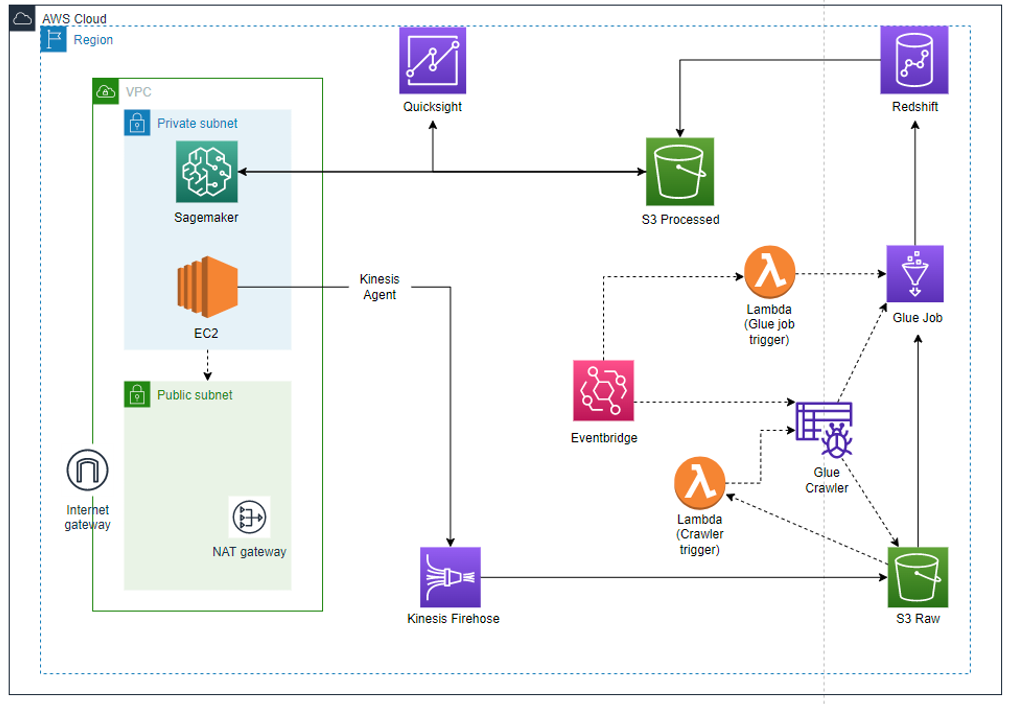
* + Customer ID
  + Gender
  + Current Date – The date at which the ride is booked
  + Current Time – The time at which the ride is booked
  + Pickup Address – The address for pickup
  + Pickup Time – The time at which the pickup is expected
  + Drop Time – Expected drop time
  + Drop Location – The destination address
  + Cost – The fare for the trip

## Proposed algorithm and Solution Steps

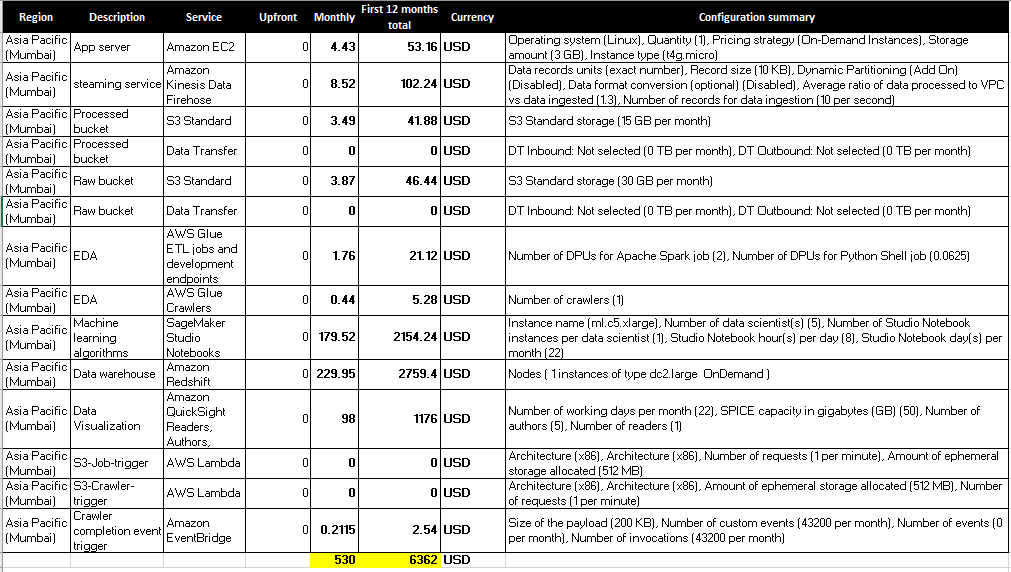
**High level Block diagram**



**Aws architecture for our solution**



## Cost Estimate of the above proposed AWS solution



Link: <https://calculator.aws/#/estimate?id=2665fe39e4b0f80ebba3d33846d6c9c1e062e90d>

## Implementation flow:

1. Move python script which is generating log files from S3 bucket to EC2 instance.

Text

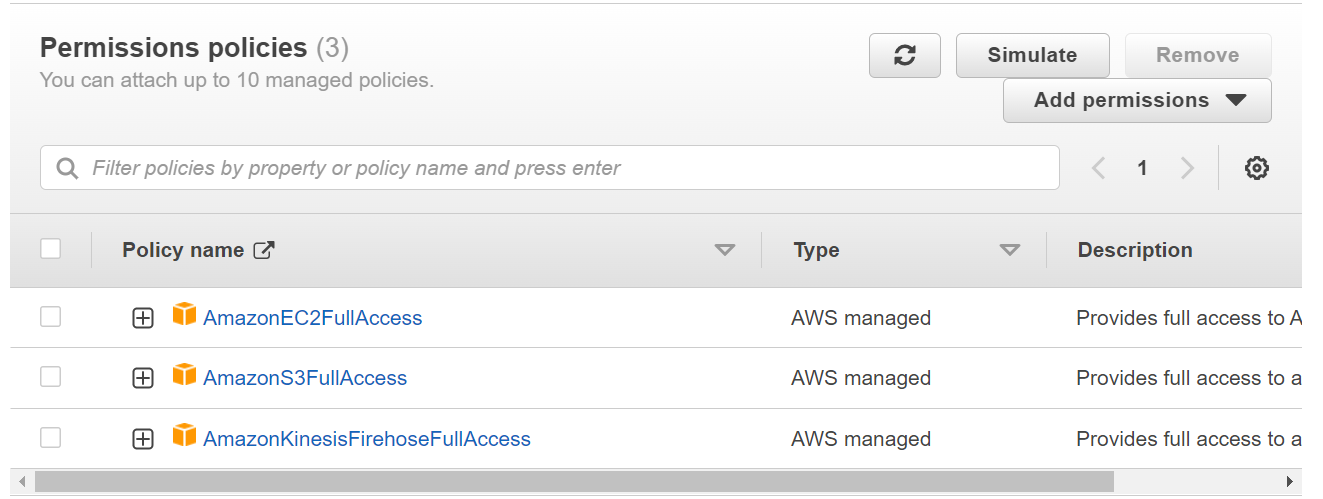
Description automatically generated

1. Install kinesis agent and then configure agent.json.

**A screenshot of a computer

Description automatically generated with medium confidence**

1. Attach necessary IAM roles with necessary permissions to EC2 (S3readOnly & FirehoseWriteOnly), Kinesis Firehose.

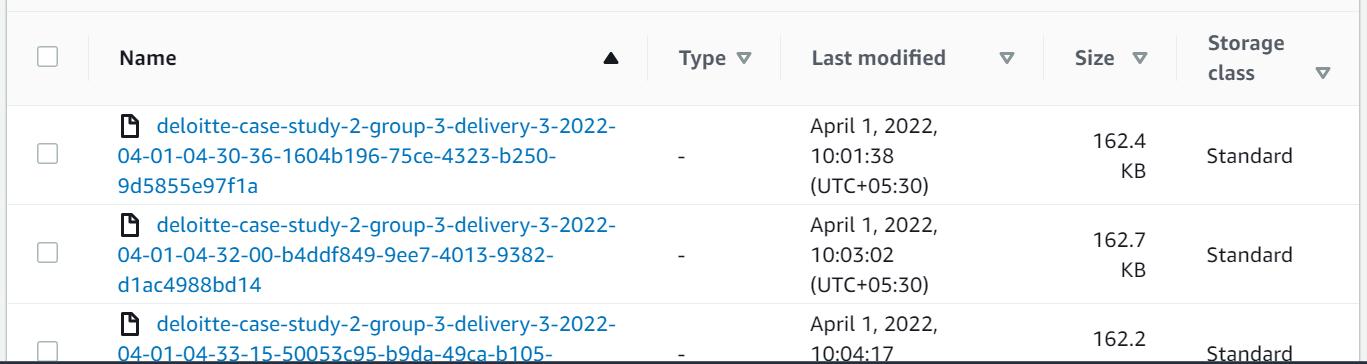


1. Execute python script on EC2 to generate log files continuously, simulating live data stream.

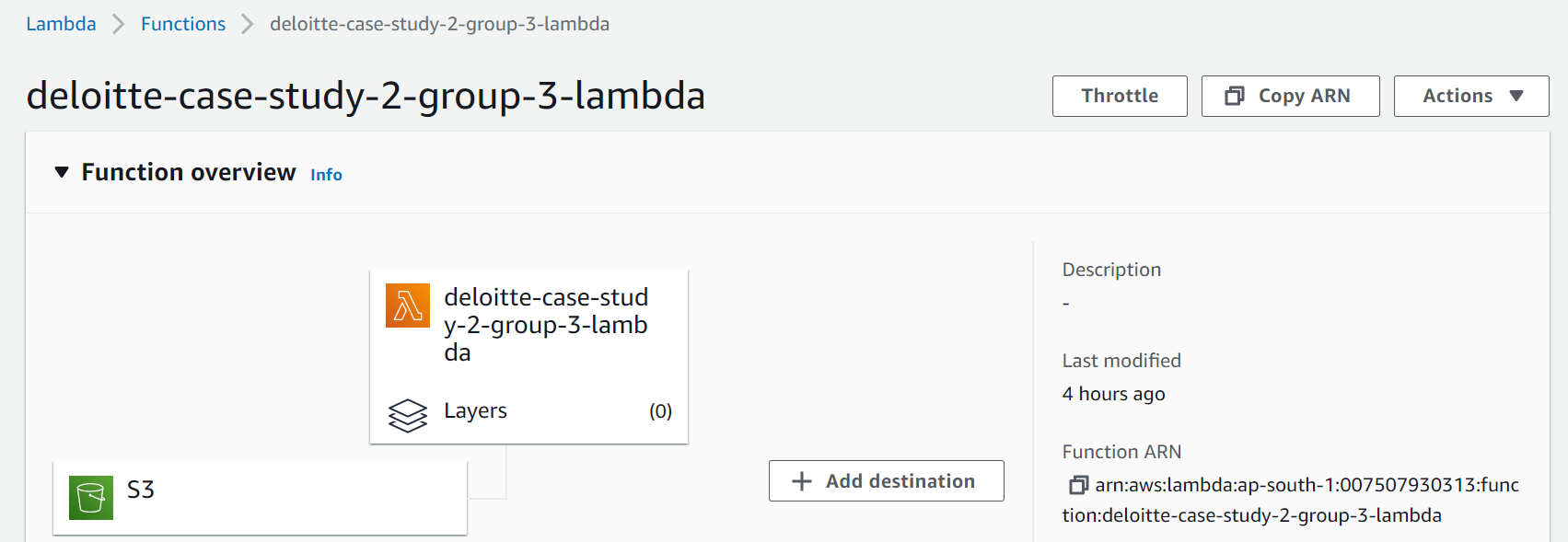
**Text

Description automatically generated**

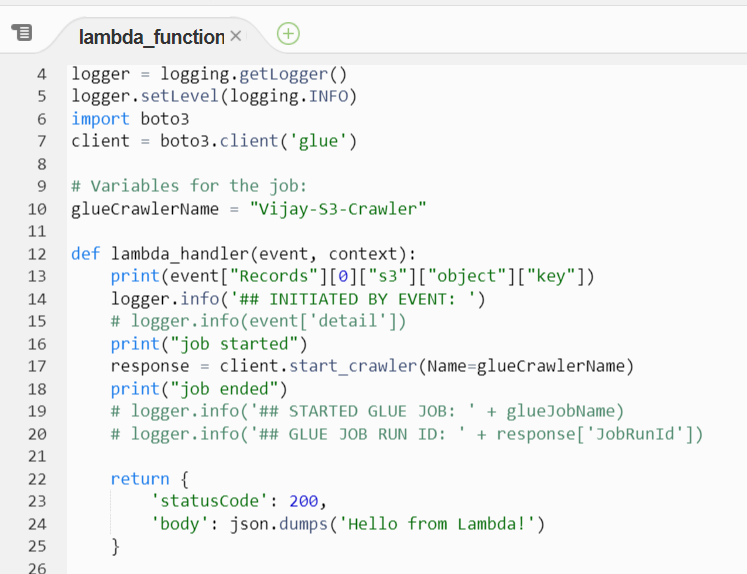
1. Generated log files transferred to S3 bucket from EC2 via Firehose with the help of kinesis agent.



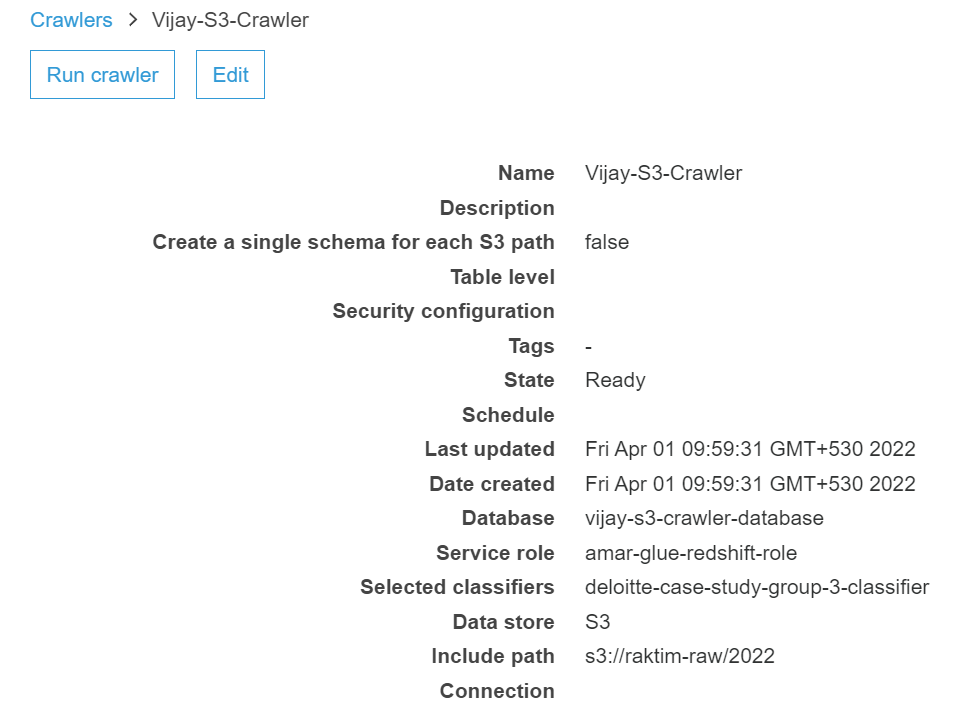
1. Configure lambda in a way that when a dataset is added in s3 the lambda should trigger crawler work.



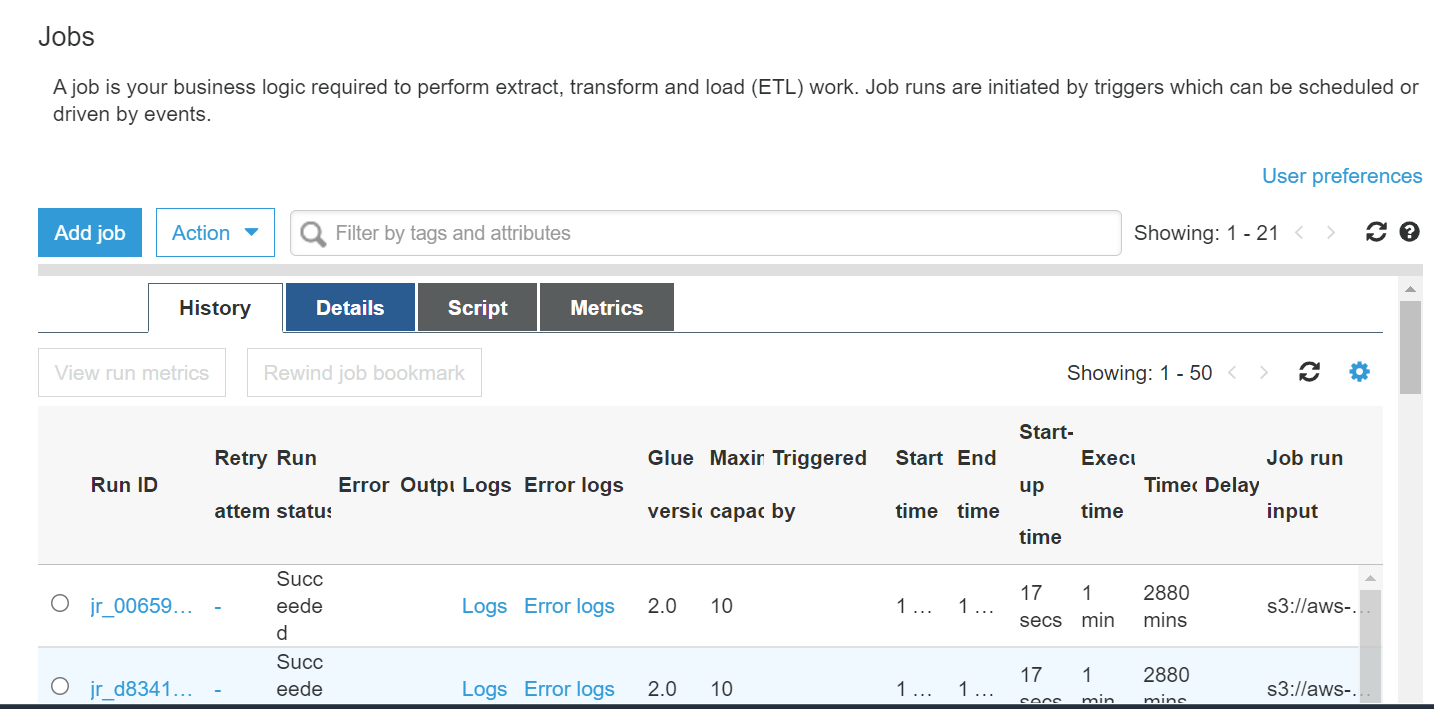
1. Code for the lambda function for whenever new log file is saved into s3 will trigger the Crawler.



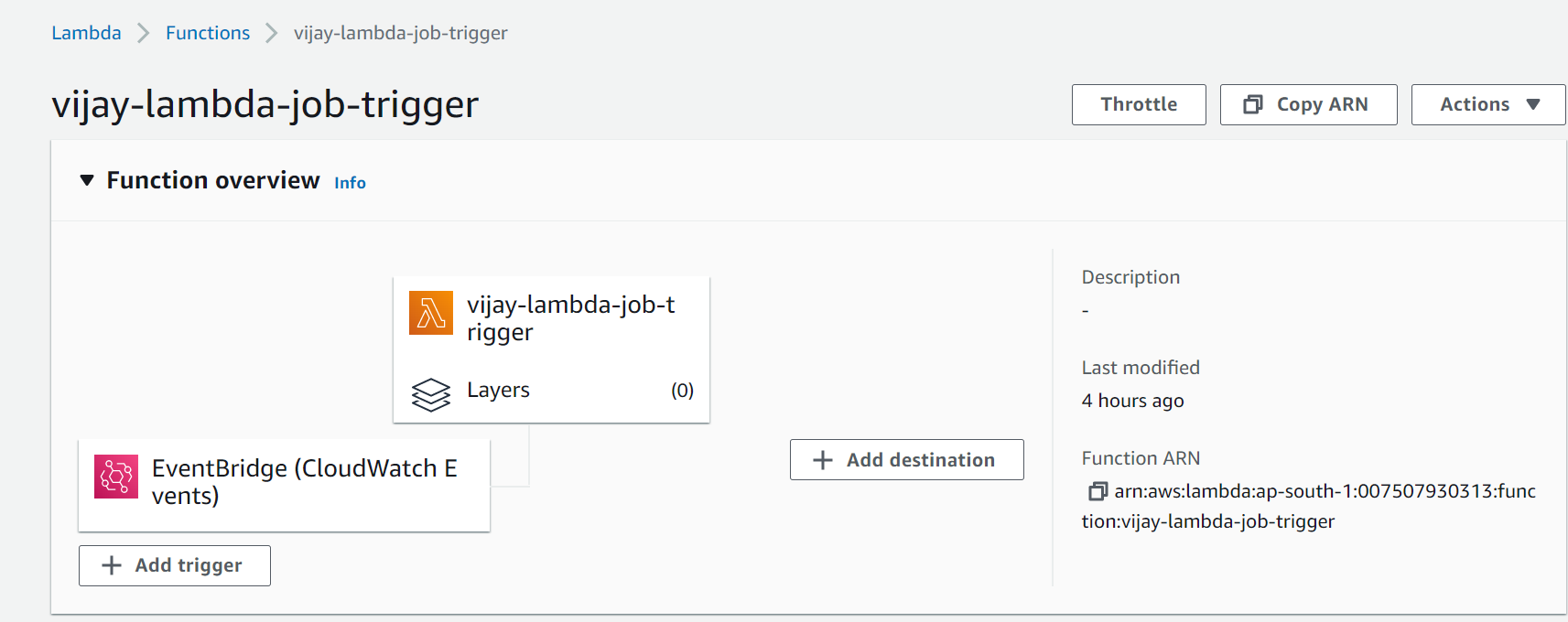
1. Crawler is created in AWS Glue to create the schema table and to transform the data.



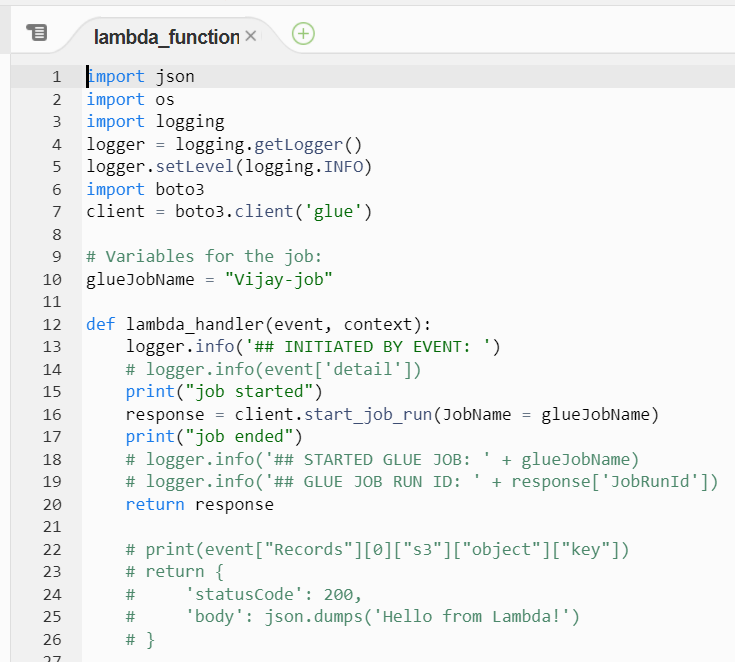
1. Glue job is executed to transform the data from S3 to Glue Crawler.



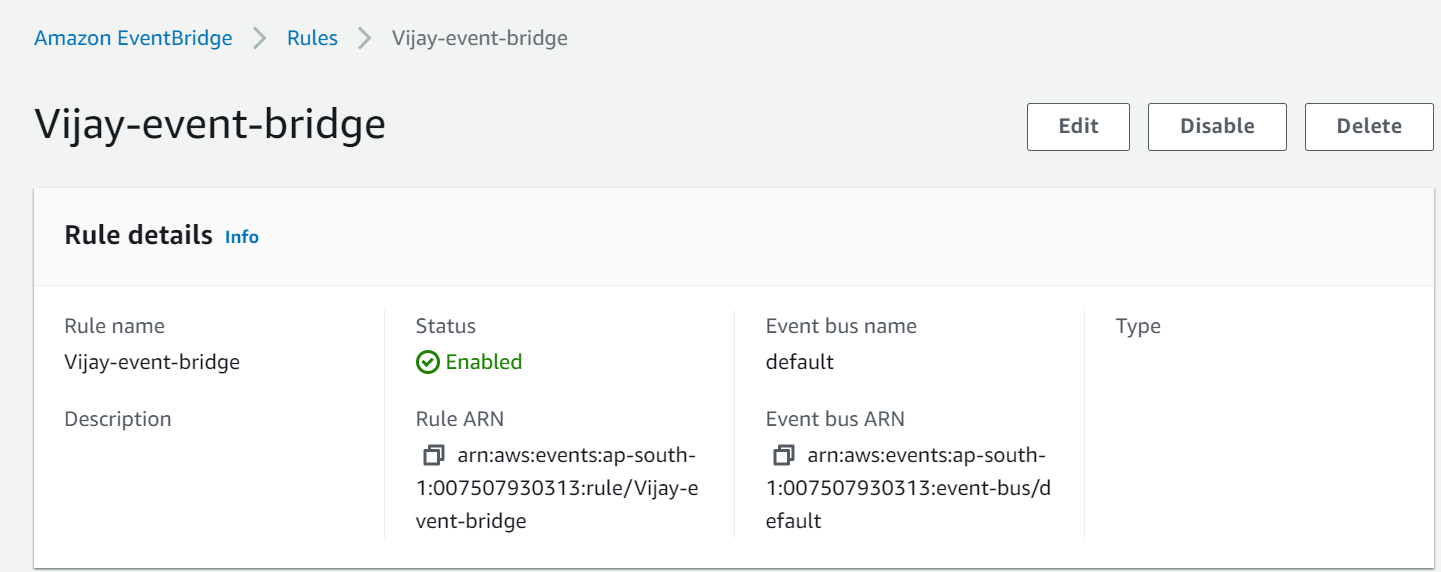
1. Once the crawler job is completed, this invokes another lambda function to insert the data into redshift tables.



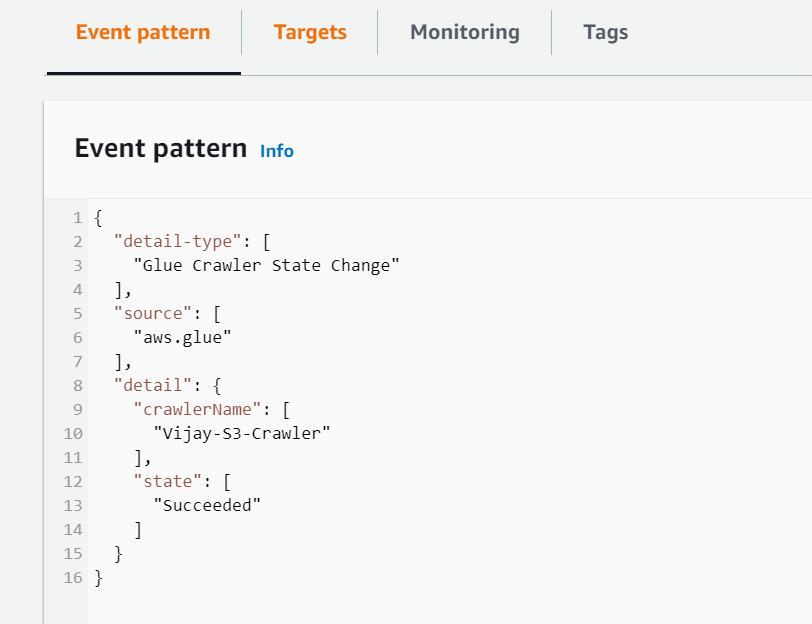
1. Code for the Lambda function



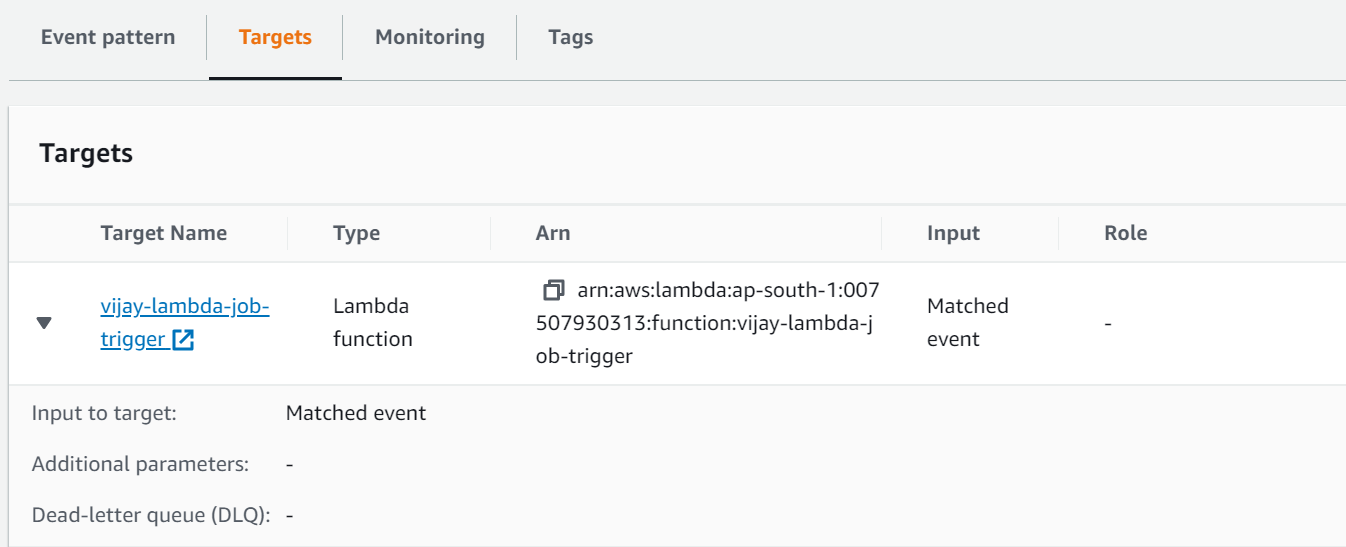
1. Create the event bridge so that to trigger the lambda once the crawler job is completed.



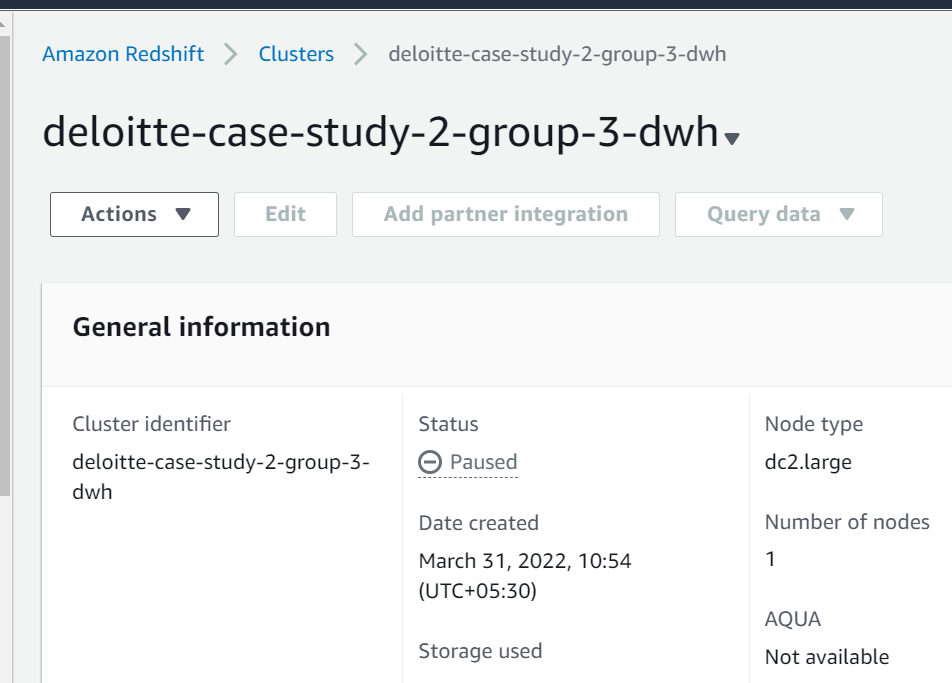
1. Below is the code for EventBridge



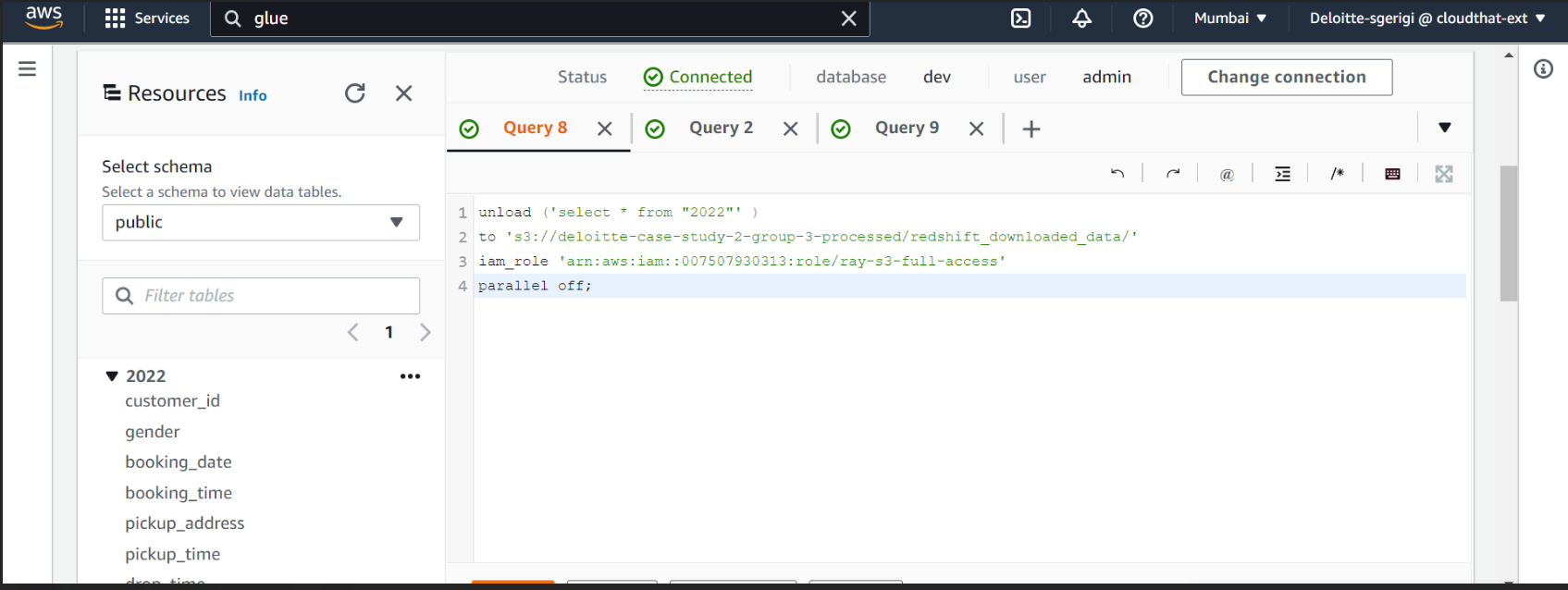
1. Below is the target that points to the Lambda.



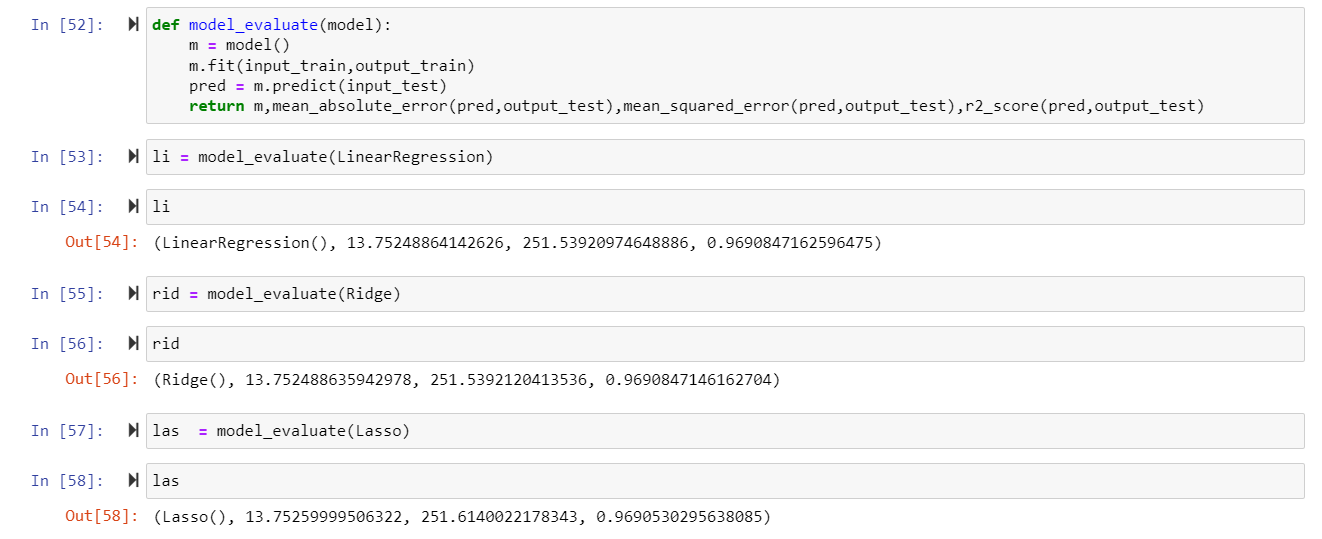
1. Create the Redshift Cluster



1. Unload the data to S3 bucket with the below query.

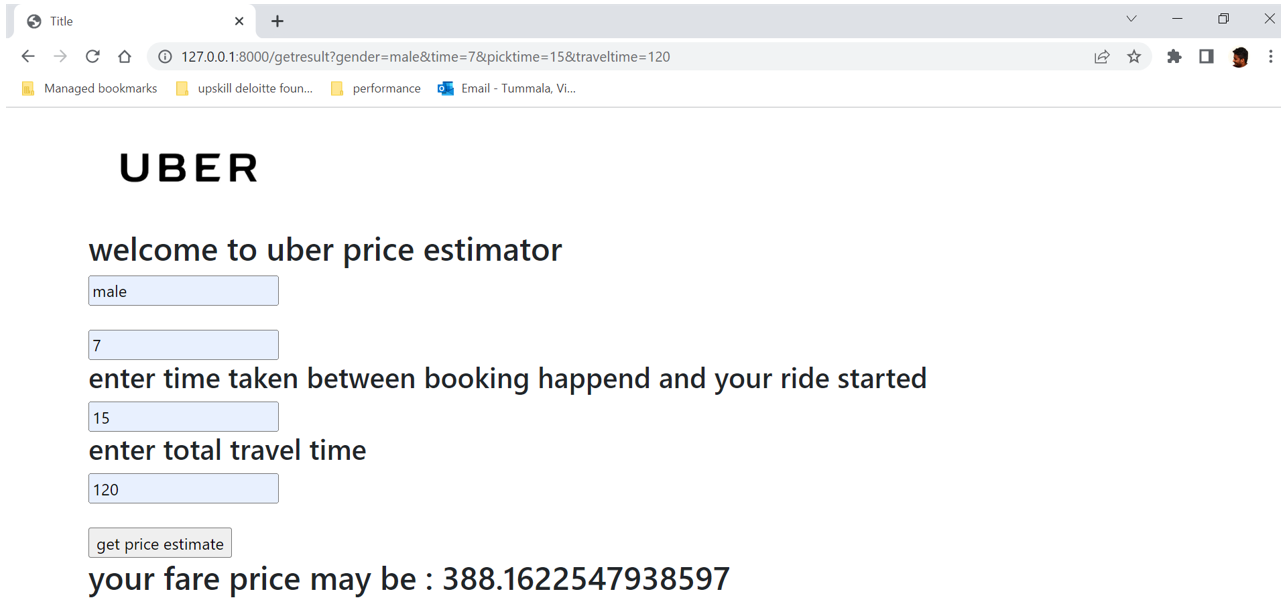


1. Removed the null values and applied algorithm models to predict the fares.

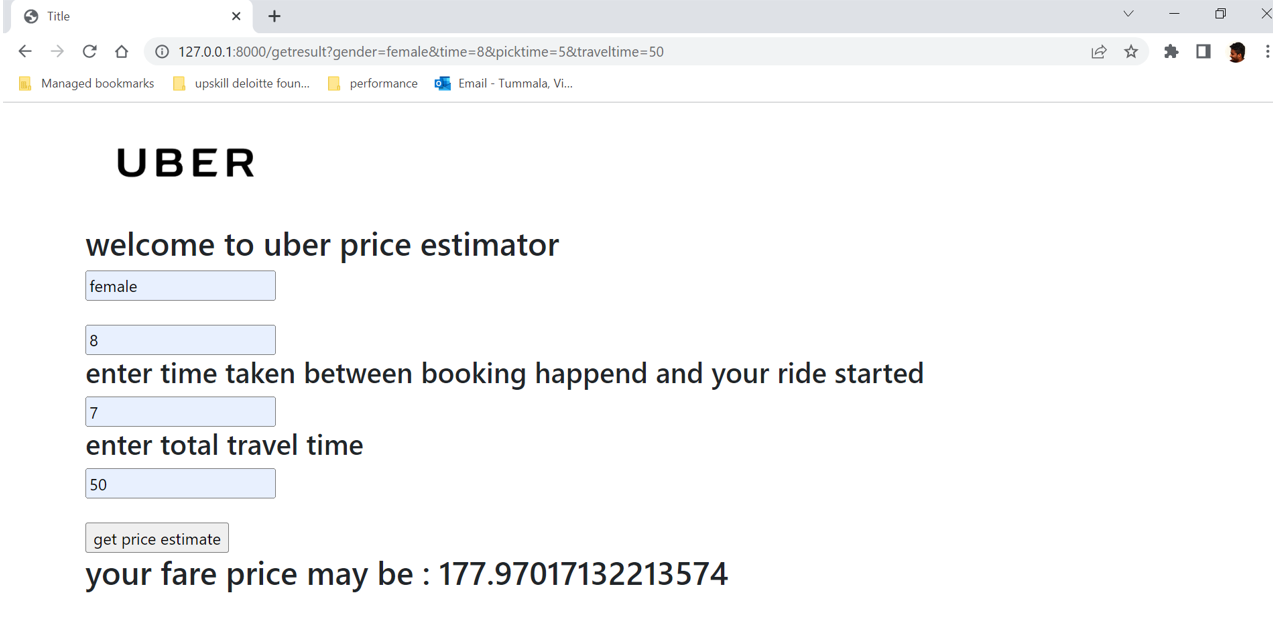


1. Created website running ML model which will predict the fare price.

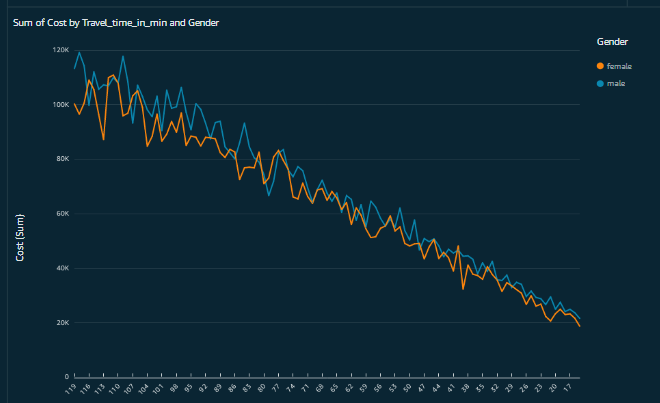
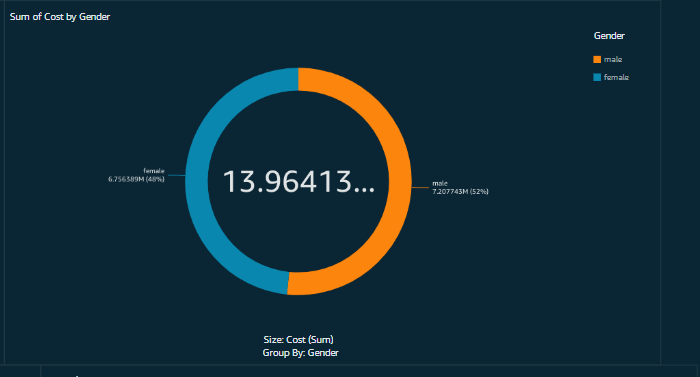
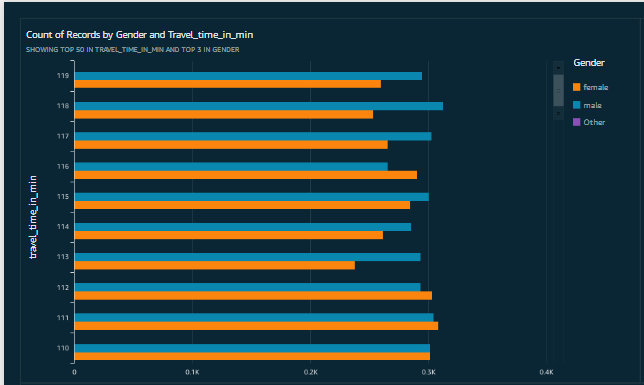
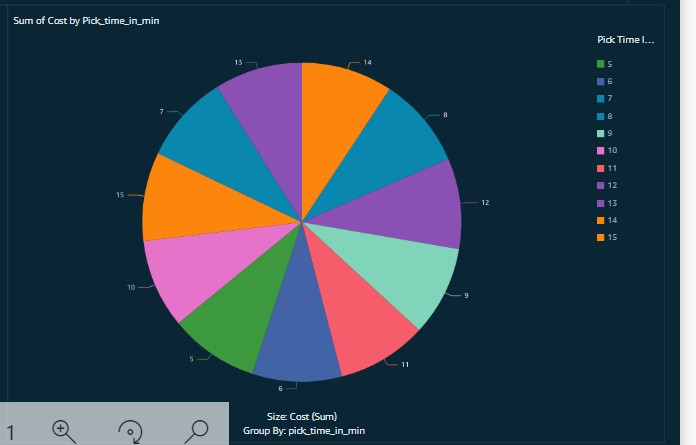
**Output1:**



**Output2:**



1. Quicksight Dashboards

## Conditions affecting Accuracy

The accuracy is taken by the machine learning models and, in the end, we can say that there is a 91% accuracy model crated which will help the uber to fare prices of the rides.

## Closure

Finally, we would like to thank the team to provide us with the opportunity to work on the data sets and create a complete solution for the problem statement.