Sample Trip\_analysis GCP console output

A black screen with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

A black screen with white text

Description automatically generated

Output files

A computer screen with white text

Description automatically generated

Output files in GCP:

<https://console.cloud.google.com/storage/browser/nyc-bucket-siva/output/avg_duration_distance_by_day;tab=objects?project=spark-taxi-analysis-424119&pageState=(%22StorageObjectListTable%22:(%22f%22:%22%255B%255D%22))&prefix=&forceOnObjectsSortingFiltering=false>

A screenshot of a computer

Description automatically generated

Copying output files to local directory  
  
A black and white image of a computer screen

Description automatically generated with medium confidence

Analysing the output generated :

A screenshot of a computer program

Description automatically generated

Similarly the output of demand\_prediction from console

A black screen with white text

Description automatically generated

From the Assignment I was able to:

**Deploying a Spark Cluster on GCP:** I have successfully deployed a Spark cluster on Google Cloud Platform (GCP) and executed Spark jobs.

**Using the Spark DataFrame API:** I have demonstrated the use of the Spark DataFrame API in my code. I've loaded data, performed transformations, aggregations, and trained machine learning models using DataFrame operations.

**Layout of Code in Spark Jobs:** My code is organized into separate Python files (tip\_analysis.py, fare\_analysis.py, traffic\_analysis.py, demand\_prediction.py) for different analysis tasks. Each file contains functions encapsulating specific tasks, and I've utilized Spark's distributed processing capabilities effectively.

**Extracting Business Value:** I have conducted various analyses using Spark, such as tip analysis, fare analysis, traffic analysis, and demand prediction. These analyses can provide valuable insights into various aspects of taxi operations in NYC, such as customer behavior, traffic patterns, fare trends, and demand forecasting. These insights can be leveraged to optimize operations, improve services, and make data-driven decisions that drive business value.