One Page Project Summary

Pavan Kalyan Imadabathini

Anushya Sivanupandian

Aasritha Devi Surapaneni

Geospatial Analysis Using Apache Spark

Group Name: Team Devils

IFT 512: Advanced Analytics Big Data/AI

Prof. Robert Rucker

Nov 26th, 2023

1.**Project Title**:-

Geospatial Analysis Using Apache Spark

2.**Team Members**: -

Aasritha Devi Surapaneni, Anushya Sivanupandian , Pavan Kalyan Imadabathini

3.**Reason for Project:-**

This project is designed to leverage the power of geospatial analysis for understanding and optimizing urban transportation systems, specifically by analyzing New York City's Yellow Cab taxi trip data. The initiative stems from the need to enhance urban mobility, reduce traffic congestion, and improve transportation services using data-driven insights. The team is driven by a passion for big data technologies and their potential to transform city planning and transportation management.

4.**Questions to be Answered:-**

- How can analyzing taxi trip data inform better urban planning and traffic management strategies?
- What insights can hotspot analysis provide to optimize taxi service allocation and routes?
- How can cluster analysis of taxi trip data help in identifying and mitigating traffic congestion hotspots?
- What business intelligence can be extracted from the data to give taxi companies a competitive edge?
- How can the findings enhance the efficiency and service quality of taxi operations in urban areas?

5.**Technologies Employed:-**

Apache Spark,Hadoop Distributed File System ,Scala ,Amazon Elastic Compute Cloud (EC2)

6.**Data structure used:-**

Arrays,DataFrames,Scalars

7.**Conclusions:-**

The project aims to provide deep insights into urban transportation patterns, revealing critical aspects of taxi service operations and urban mobility. These insights are expected to be instrumental in guiding city planners, policymakers, and taxi service providers in making data-driven decisions. The flexibility of the methods and technologies used in this project also allows for scalability and adaptability to other similar urban datasets and geographies.