A GEOSPATIAL BIG DATA ANALYSIS OF NYC TAXI TRIP

AIFAA FARZANAH BINTI SAIFULLIZAN

AMINAH UMAIRAH BINTI AZLI

Final Year Project Report Submitted in
Partial Fulfilment of the Requirements for the
Diploma in Geospatial Technology
in the College of Built Environment
Universiti Teknologi MARA

JULY 2024

ABSTRACT

A Geospatial Big Data Analysis of NYC Taxi Trip

New York City (NYC) is a major metropolis globally, with one of the busiest and most intricate public transport systems. Taxis are an indispensable means of quick and adaptable transit during New York City's chaos, covering routes that the subway and bus systems would not be able to reach. This project will analyse the taxi trip data using the analysis tools in the ArcGIS Pro software. This data will also be visualised using Power BI Desktop for the purpose of visualizing the results. This project aims to determine the NYC taxi pick up and drop off locations based on one month data. There are several problem statements that justify the creation of this project. The first one is the problem of crowding people waiting for taxis may not happen. So, with this project it can help society in New York City. The second project has benefited to the tourists who come to New York City because it can help them to get taxis and know the hotspots where taxis always pick up and drop off. This project contains three (3) objectives which are to manage NYC Taxi Trip big data for data analysis. The data will be managed using Microsoft Excel according to what data is required. The second objective is to determine taxi pickup and drop off hotspots using geospatial analysis. The last objective is to apply data visualization of the NYC Taxi Trip analysis result where Microsoft Power BI Desktop will be used to create data analysis. The study revealed that the most pick-up and drop off location is zones in Manhattan. This project will help society and tourists in New York City to avoid the problem of waiting longer for taxis because they can know the hotspot areas where taxis always pickup and drop-off their passengers. This research contributes to the field of transportation by providing a dashboard that has information about the data used to unravel existing problems or make an analysis using the dashboard.

ACKNOWLEDGEMENTS

We would like to sincerely thank Sr. Gs. Azlizan Adila binti Mohamad, our supervisor, for her important advice and assistance during this project. Their knowledge and support have greatly influenced the direction of our work. We also want to express our gratitude to our friends and students for being so welcoming and for creating a lively atmosphere that encouraged cooperation and conversation. I am grateful to the Seri Iskandar campus of UiTM Perak branch for lending me the tools and assistance I needed to complete my research. I sincerely appreciate your assistance. Ultimately, I express my gratitude to my family for their consistent support and motivation. Our main source of inspiration has been their belief in me.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

1.1 Background of Study

New York City (NYC) is a major metropolis globally, with one of the busiest and most intricate public transport systems. Taxis are an indispensable means of quick and adaptable transit during New York City's chaos, covering routes that the subway and bus systems would not be able to reach. Optimising operations and boosting profitability in the taxi sector requires a thorough awareness of regional preferences, client demand patterns, and factors impacting customer satisfaction, such tipping. It is a difficult task in today's fiercely competitive digital environment to increase the profitability of taxi services through the analysis of client behaviour and preferences.

Geospatial big data refers to a vast amount of data that contains geographic information. This data is typically generated from various sources such as satellites, GPS devices, social media check-ins, sensors, and other location-aware technologies. It includes information like coordinates, addresses, boundaries, and other spatial attributes. The significance of geospatial big