NYC Taxi & Citi Bike Data Analysis
Project Plan

Version 1.0

Table of Contents

1. Introduction

1.1 Purpose of this document

The purpose of this document is to provide a detailed plan for our project titled 'NYC Taxi & Bike Data Analysis', which has been taken up to learn, research and understand different aspects of big data technologies and applications. This specific project has been chosen to uncover useful facts and meaningful insights about New York Taxi and Citi bike data, COVID-19 impact on the businesses, helping drivers with pickup hot spots, comparison between taxi & bike business etc. This document includes details about organization, roles, deliverables, project risks, time plans and financial plans.

1.2 Intended Audience

This document shall be used in all phases of the project as a guideline. Intended audience of this project plan document will be our team - Abinaya Seshadre, Neelima Jagtap, Rishikumar Ravichandran, Sruthi Mallarapu and our guide, Professor Andrew Bond.

1.3 Scope

This document defines the project plan of the NYC Taxi & Bike Data Analysis application. The overview includes objectives of the project, organization of the project team, development process that is going to be used during the project, assessment of possible risks, communication used between project stakeholders and project plan that includes time schedule and activity plan.

1.4 Definitions and acronyms

1.4.1 Definitions

Keyword	Definitions
NYC Taxi & Bike Data Analysis	The name of the project
Project Supervisor	A person in charge of supervising the project
Project Leader	A person in charge of organizing the team and communicating with the project supervisor
Team Member	An active member of the team responsible for making the job done
Milestone	A time in a project that marks the end of a project phase or the completion of an important deliverable.
Git	Version control system that will be used in this project
Scrum	An iterative and incremental agile software development method for managing software projects and product or application development
Scrum sprint	The basic unit of development in Scrum
Scrum master	Ensures the smooth working of the Scrum team and enforces Scrum practices

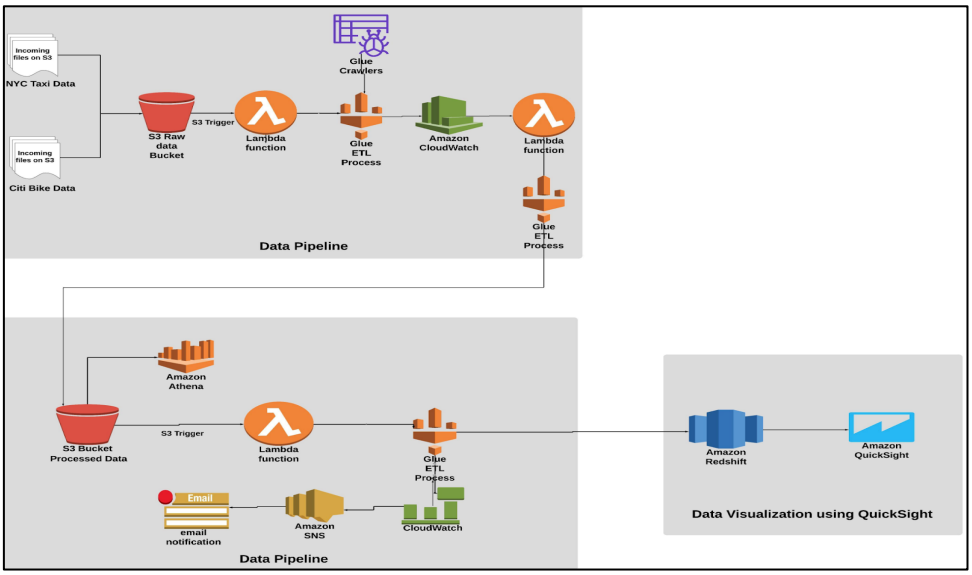
1.5 References

1. http://www.scrum.org/

2. Background and Objectives

As COVID 19 pandemic affected most of the business sectors, it also affected on daily transport sector. More specifically the drivers who is dependent on online taxi services like New York Taxi, Uber etc. are suffering because of this pandemic situation. In this project, by analyzing New York taxi and bike data, the outcome can help New York taxi drivers.

3. Architecture & High-Level Design



4. Organization

4.1 Project group

First Name	Last Name	Responsibility (roles)
Abinaya	Seshadre	Data ingestion, data pipeline, documentation
Neelima	Jagtap	Dataset, data visualization, documentation
Rishikumar	Ravichandran	Dataset, data visualization, documentation
Sruthi	Mallarapu	Data ingestion, data pipeline, documentation

5. Development process

The project will use following tools.

Purpose	Tool/Process
Project Meetings	Zoom meetings
Project Proposal	MS Word
Data Sources	NYC TLC Taxi datasets, NYC Citi Bike datasets
Data Model	Lucid Chart
Data Pipeline	Amazon S3, Glue,Lambda, CloudWatch, SNS,Redshift,Quicksight
Data Cleansing	AWS Glue
Data Store	S3, AWS Redshift
Data Analysis	Redshift SQL queries
Visualization	AWS QuickSight
Project Presentation	MS PowerPoint
Project Report	MS Word
Code Version Control	GitHub

6. Deliverables

Deliverable Name	Planned week	Delivered week	Notes
Project Abstract	Week 1	Week 1	Finalized dataset
Project Design	Week 2	Week 2	Decided which AWS services to use
Data Model	Week 3	Week 3	Finalized the data model for project
Data Pipeline	Week 4	Week 5	Data pipeline formation in AWS
Data Cleansing	Week 5	Week 6	Data cleansing using ETL
Data Analysis	Week 6	Week 7	Performed in depth data analysis
Data Visualization	Week 7	Week 8	Dashboard building of analyzed data using Quick Sight
Project Testing	Week 8	Week 9	Tested imported data, scripts, pipelines, dashboards
Report Draft	Week 9	Week 10	Divided the documentation work
Report Final	Week 9	Week 10	Developed final report after edits

7. Project Challenges

Challenges	Resolutions
AWS is new	Classes, ISA, professor
Implementation issues	CloudWatch logs, online blogs, team, peers

8. Communication

8.1 Collaboration

- Project Planning (grooming):
  - 3 Hours meeting to discuss overall project planning and task division between team members.
- Daily Scrum:
  - Monday to Friday daily standup for 15 minutes to discuss what progress made and any blocking issues.
- Documentation & Demo preparation:
  - One-week daily meeting (30 minutes) for project documentation and demo preparation.

8.2 Git

All source code and finished documentation will be uploaded to Github repository.

https://github.com/sruthimallarapu/Dataluation_Data228_Project

9. Project plan

9.1 Time schedule

Id	Milestone Description	Responsible Dept./Initials	Finished week	Forecast Week	+/-	Actual	Metr.	Rem.	
	Finalize Project Topic	ALL	Week 1	Week 1					
	Figure out Data sources	ALL	Week 1	Week 1					
	Prepare an abstract	Abhinaya	Week 1	Week 1					
	Design Data Model	Rishi	Week 2	Week 2					
	Architect Data pipeline	Sruthi	Week 3	Week 2					
	Data collection and cleansing	Neelima	Week 4	Week 3					
	Implement ETL job	Sruthi	Week 5	Week 5					
	Analyze Data	Rishi	Week 7	Week 6					
	Build visualization dashboards	Rishi, Neelima	Week 8	Week 8					
	Project Testing	Abhinaya	Week 9	Week 9					
	Prepare Demo	Abhinaya	Week 9	Week 9					
	Prepare Draft report	Abhinaya	Week 10	Week 10					
	Prepare Final report	ALL	Week 10	Week 10					

9.1.1 Remarks

Remark Id	Description
	Everyone contributed well and delivered the expected results in time as per the plan.

9.2 Test plan

Test No.	001	**Phase:**	1	**Author:**	<username>	Date:
Test Category:						
Software Product:	AWS glue					
Test Title:	Test if glue job is producing the correct prepared data					
Test Purpose:	Testing the ETL job					
Test Setup:	AWS glue, Lambda functions, Raw data S3 buckets					
Prerequisites:						
Procedure:	Upload a raw data file into S3 bucket and observe the execution of ETL job					
Checks:	**Lambda execution, Crawler creation, Processed Data generation, Notification message generation.**					
Expected Results:	Lambda functions should invoke on respective triggers. Data crawler should be created. Processed data should be generated and notification should be generated					
Result:	Results are as expected.					
Reason for Failure:	NA					
Remarks:	NA					

9.2.1 Testing Remarks

Remark Id	Description
	Test execution is successful

10. References

- https://www1.nyc.gov/site/tlc/index.page
- https://zacks.one/aws-athena-lab/
- https://docs.aws.amazon.com/code-samples/latest/catalog/code-catalog-python-example_code-kinesis.html
- https://docs.aws.amazon.com/ses/latest/DeveloperGuide/event-publishing-redshift-firehose-stream.html