**Project Charter Document**



**Project Name:** Analysis of Yelp dataset using HIVE

**Department:** Customer Review

**Product/Process:** Data Analysis



**Prepared By**

|  |  |
| --- | --- |
| **Document Owner(s)** | **Project/Organization Role** |
| Abhishek Tiwari | Futurense Technologies |
| Manjeet Raj | Futurense Technologies |
| Rushal Suvarna | Futurense Technologies |
| Shreepad Parakhi | Futurense Technologies |

**Project Charter Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Change Description** |
| 1.0 | 13/03/2022 | Shreepad | Document created |
|  |  |  |  |

**TABLE OF CONTENTS**

[1 PROJECT CHARTER PURPOSE 3](#_Toc98107474)

[2 PROJECT EXECUTIVE SUMMARY 3](#_Toc98107475)

[3 PROJECT SCOPE 3](#_Toc98107476)

[3.1 Goals and Objectives 3](#_Toc98107477)

[3.2 Project Duration 4](#_Toc98107478)

[start date: 12/03/2022 Estimated End date: 19/03/2022 4](#_Toc98107479)

[4 PROJECT CONDITIONS 5](#_Toc98107480)

[4.1 Project Assumptions 5](#_Toc98107481)

[4.2 Project Issues 5](#_Toc98107482)

[5 Project Structure Approach 5](#_Toc98107483)

[6 PROJECT REFERENCES 6](#_Toc98107484)

[7 APPROVALS 6](#_Toc98107485)

# PROJECT CHARTER PURPOSE

* Get idea about business problem and constraints.
* To get high level view of project scope.



# PROJECT EXECUTIVE SUMMARY

* Project goals
* Objectives
* Scope
* Assumptions
* Timeline
* Approach



# PROJECT SCOPE

## Goals and Objectives

|  |  |
| --- | --- |
| **Goals** | **Objectives** |
| * To perform Hive analytics on Customer Demographics data using big data tools such as Hive. | * Yelp is a community review site and an American multinational firm based in San Francisco, California. It publishes crowd-sourced reviews of local businesses as well as the online reservation service Yelp Reservations. * Yelp has made a portion of their data available in order to launch a new activity called the Yelp Dataset Challenge, which allows anyone to do research or analysis to find what insights are buried in their data. * Due to the bulk of the data, this project only selects a subset of Yelp data. User and Review dataset is considered for this session. |

## Project Duration

## start date: 12/03/2022 Estimated End date: 19/03/2022

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Milestone** | **Date Estimate** | **Deliverable(s) Included** | **Confidence Level** |
| * Project Scope Definition | [12/03/2022]  -  [13/03/2022] | * Project Scope Document (Word or PPT format) * Should include data source definition and dataset. Include the business problem and how what solution is expected to solve the problem | [High] |
| * Data Source | [12/03/2022]  -  [13/03/2022] | * To include the data source, data preparation and explanation * To include the data source, data preparation and explanation | [High] |
| * Detailed Project Architecture | [14/03/2022] | * PPT slides * To include the entire architecture | [High] |
| * Complete Code Submission-  Data Extraction and Ingestion | [15/03/2022] -  [17/03/2022] | * Code snippets in PPT format + github links * Include:   1. Data extraction and ingestion   2. Data storage and processing | [Medium] |
| * Final Deliverable – Complete Project Solution and Its Business Impact | [18/03/2022] |  | NA |
| * Project Presentation | [19/03/2022] | * Prepare final PPT | NA |

# PROJECT CONDITIONS

## Project Assumptions

* Data is in cleaned format and no need of preprocessing
* Data won’t be updated in near future.

## Project Issues

**Priority Criteria**

1 − High-priority/critical-path issue; requires immediate follow-up and resolution.

2 − Medium-priority issue; requires follow-up before completion of next project milestone.

3 − Low-priority issue; to be resolved prior to project completion.

4 − Closed issue.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Date** | **Priority** | **Owner** | **Description** | **Status & Resolution** |
| 1 | 13-03-2022 | High | Manjeet | Analyzing given yelp dataset | In Progress |
| 2 | 13-03-2022 | Medium | Rushal | Loading dataset in HDFS | In queue |
| 3 | 12-03-2022 | Medium | Abhishek | Data Mining | Closed Issue |
| 4 | 13-03-2022 | High |  | Identifying necessary Serde and jars | Not yet started |



# Project Structure Approach

* Data sourcing
* Load data in HDFS
* Load data in hive staging tables
* Apply partitions based on year and month
* Apply bucketing based on UserId
* Perform analytical queries

# PROJECT REFERENCES

|  |  |
| --- | --- |
| **Milestone** | **Reference** |
| Data Source | www.yelp.com/dataset/download/ |
| A Project Approach | SmidsyTechnologies |
| Da Data Documentation | <https://www.yelp.com/dataset/documentation/main> |



# APPROVALS

**Prepared by** Shreepad Parakhi\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Approved by** Mr. Saif\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Mentor

