California Housing Market Application

Project Plan

Version 1.0

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Description** | **Author** |
| NA |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table of Contents

[1. Introduction 4](#_Toc119933611)

[1.1 Purpose of this document 4](#_Toc119933612)

[1.2 Intended Audience 4](#_Toc119933613)

[1.3 Scope 4](#_Toc119933614)

[1.4 Definitions and acronyms 4](#_Toc119933615)

[1.4.1 Definitions 4](#_Toc119933616)

[1.4.2 Acronyms and abbreviations 6](#_Toc119933617)

[1.5 References 6](#_Toc119933618)

[2. Background and Objectives 6](#_Toc119933619)

[3. Architecture & High Level Design 6](#_Toc119933620)

[4. Organization 7](#_Toc119933621)

[*4.1* Project group 7](#_Toc119933622)

[4.2 Customer 8](#_Toc119933623)

[5. Development process 8](#_Toc119933624)

[*6.* Deliverables 8](#_Toc119933625)

[7. Project risks 8](#_Toc119933626)

[8. Communication 10](#_Toc119933627)

[8.1 Collaboration 10](#_Toc119933628)

[8.2 Git 10](#_Toc119933629)

[9. Project plan 10](#_Toc119933630)

[9.1 Time schedule 10](#_Toc119933631)

[9.1.1 Remarks 11](#_Toc119933632)

[9.2 Test plan 12](#_Toc119933633)

[9.2.1 Testing Remarks 13](#_Toc119933634)

[10. References 13](#_Toc119933635)

# Introduction

## Purpose of this document

The purpose of this document is to provide a detailed description of the project ‘California Housing Market Application’ which is designed to analyze the impact of the presence of school districts on house prices and predict the best house price for the home buyers. This document includes details about organization, roles, deliverables, project risks, time plans and financial plans.

## Intended Audience

This document shall be used in all phases of the project as a guideline. Intended audiences of this project are all project stakeholders:

* Project supervisor
* Project leader
* Team members
* Home buyers

## Scope

This document defines the project plan of the ‘California Housing Market 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.

## Definitions and acronyms

### Definitions

|  |  |
| --- | --- |
| **Keyword** | **Definitions** |
| Project Name | California Housing market Application |
| Project Supervisor | Andrew Bond |
| Project Leader | Sree Divya Cheerla |
| Team Member | Nupur Pathak  Revathi Boopathi  Sree Divya Cheerla  Vani Bhat |
| Milestone | A time in a project that marks the end of a project phase or the completion of an important deliverable. |
| Git | <https://github.com/DivyaCheerla> |
| Scrum | An iterative and incremental agile software development method for managing software projects and product or application development |
| ClickUp | Web-based tool for integrated agile project management and collaboration based on Scrum |
| Scrum sprint | The basic unit of development in Scrum |
| Scrum master | Ensures the smooth working of the Scrum team and enforces Scrum practices |
| Product owner | Responsible for product management and its quality |

### Acronyms and abbreviations

|  |  |
| --- | --- |
| **Acronym or**  **abbreviation** | **Definitions** |
| AWS | Amazon Web Services |
| ETL | Extract, Transform and Load |
|  |  |
|  |  |
|  |  |

## References

1. <https://www.zillow.com/>
2. <https://www.ccsa.org/what-we-do/student-success>

# Background and Objectives

The surge in demand for the US housing market and shortage of inventory has made it difficult for people to find their dream home with all the features they desire with ease. Hybrid work culture resulting from the pandemic is fueling the demand in the housing market. There are various parameters that are impacting the property values such as school proximity, migration of people impacting the population density, access to public transportation, and others.

The main objective of this project is to build a California housing market application which predicts the house price for the home buyers. It also shows the impact on house prices due to the presence of renowned schools which can assist home buyers make the right decision.

# Architecture & High-Level Design

Diagram

Description automatically generated

# Organization

MSDA DATA -228- Big Data Technology and Applications

## Project group

|  |  |  |
| --- | --- | --- |
| **Name** | **Initials** | **Responsibility (roles)** |
| Nupur Pathak | NP | Web scraping (Python), Tableau public |
| Revathi Boopathi | RP | AWS S3, AWS SageMaker |
| Sree Divya Cheerla | DC | AWS Athena, Website, Tableau Desktop |
| Vani Bhat | VB | AWS Glue, AWS Cloud Formation |

## Customer

The target customers are listed below:

* Home buyers
* Home sellers

# Development process

The project has made use of below tools and technologies starting from design to end-product.

|  |  |
| --- | --- |
| **Tools and Technologies** | **Usage** |
| Python, Jupyter Notebook  (Web Scraper) | To scrape house data from Zillow |
| AWS S3 | Bucketing and Data Storage |
| AWS CloudFormation | Load IAM roles, load database and tables, map S3 path to fetch data |
| AWS Glue | Database, Tables, ETL (Extract, Transform and Load) |
| AWS SageMaker | Implement Linear Regression Machine Learning Model |
| AWS Athena | Query the data in AWS GLUE Tables and connect to Tableau |
| Tableau desktop | Dashboard Creation |
| Wixite Website | Frontend UI |

# Deliverables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No** | **Output** | **Planned week** | **Promised week** | **Late +/-** | **Delivered week** |
| **1** | Abstract | Sept 6 | Sept 6 | 0 | Sept 6 |
| **2** | Design | Sept 12 | Sept 12 | 0 | Sept 12 |
| **3** | Coding | Sept 28 | Sept 28 | 0 | Sept 28 |
| **4** | Testing | Oct 24 | Oct 24 | 0 | Oct 24 |
| **5** | Documentation | Nov 20 | Nov 20 | 1 | Nov 21 |

# Project risks

|  |  |  |
| --- | --- | --- |
| **Possibility** | **Risk** | **Preventive action** |
| NA |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Communication

* Zoom
* Scrum meetings

## Collaboration

* Goggle drive
* Git

## Git

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

Repository URL: <https://github.com/DivyaCheerla>

# Project plan

## Time schedule

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Id** | **Milestone**  **Description** | **Responsible Dept./Initials** | **Finished week**  **plan** | **Actual week** |
| 1 | Abstract | Team | Sept 6 | Sept 6 |
| 2 | Architecture design | Team | Sept 12 | Sept 20 |
| 3 | Data Collection | Team | Sept 21 | Sept 25 |
| 4 | AWS S3 bucketing and versioning | Vani | Sept 25 | Sept 28 |
| 5 | AWS Cloud Formation analysis and load resources | Vani/Revathi | Oct 1 | Oct 5 |
| 6 | AWS Glue- Data Transformation Coding | Vani/Revathi | Oct 10 | Oct 15 |
| 7 | Glue Crawlers job creation | Divya | Oct 15 | Oct 20 |
| 8 | AWS SageMaker – Linear Regressor modelling | Vani/Revathi | Oct 25 | Oct 30 |
| 9 | AWS Athena – access S3 data | Divya, Nupur | Nov 1 | Nov 15 |
| 10 | Tableau public dashboard | Nupur | Nov 11 | Nov 19 |
| 11 | Website | Divya | Nov 11 | Nov 20 |
| 12 | Documentation | Team | Nov 14 | Nov 21 |
|  |  |  |  |  |
|  |  |  |  |  |

### Remarks

|  |  |
| --- | --- |
| **Remark Id** | **Description** |
|  | NA |
|  |  |
|  |  |
|  |  |

## Test plan

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 001 | **Phase:** | 1 | **Author:** | Vani Bhat | Date: 10/14/2023 |
| **Test Category:** | | AWS Data Transformation | | |  |  |
| **Software Product:** | | AWS Glue visual editor | | | |  |
| **Test Title:** | | Data Transformation using AWS Glue | | | | |
| **Test Purpose:** | | Verify if the .csv is updated after applying transformation like changing datatypes, filter, dropping duplicates, handling null values | | | | |
| **Test Setup:** | | Load .csv file in AWS S3  Create an instance of GLUE Visual editor | | | | |
| **Prerequisites:** | | AWS Could Formation has fetched the AWS S3 path successfully | | | | |
| **Procedure:** | | 1. Create visual in the instance of GLUE Visual editor  2. Apply change schema  3. Apply drop duplicates  4. Apply filter for state = “CA”  5. Apply custom SQL Query to handle “#N/A” | | | | |
| **Checks:** | | Verify if the given data transformation was successfully done on the input .csv file from S3 | | | | |
| **Expected Results:** | | 1. Data types of the variables are updated  2. Duplicates are dropped  3. Records are filtered with CA  4. #N/A values are excluded | | | | |
| **Result:** | | 1. Data types of the variables are updated  2. Duplicates are dropped  3. Records are filtered with CA  4. #N/A values are excluded | | | | |
| **Reason for Failure:** | | NA | | | | |
| **Remarks:** | | Test case passed | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test No.** | 002 | **Phase:** | 1 | **Author:** | Nupur Pathak | Date: 11/15/2023 |
| **Test Category:** | | External interface testing | | |  |  |
| **Software Product:** | | AWS Athena and Tableau dashboard | | | |  |
| **Test Title:** | | Connectivity of AWS to Tableau and visualizations | | | | |
| **Test Purpose:** | | Verify if the conversion has established properly between AWS and Tableau to visualize the data | | | | |
| **Test Setup:** | | Load data in tableau from AWS Athena | | | | |
| **Prerequisites:** | | Installation of JDBC driver | | | | |
| **Procedure:** | | 1.Launch AWS Athena  2.Query the data from data catalog in AWS GLUE  3. Save the view to S3  4. Load the data to tableau from the view crated in AWS Athena | | | | |
| **Checks:** | | Verify if Schema and data is loaded into Tableau | | | | |
| **Expected Results:** | | Data from AWS should be loaded into Tableau appropriately | | | | |
| **Result:** | | Data along with proper schema is loaded to Tableau | | | | |
| **Reason for Failure:** | | NA | | | | |
| **Remarks:** | | Test case passed | | | | |

### Testing Remarks

|  |  |
| --- | --- |
| **Remark Id** | **Description** |
|  |  |
|  |  |
|  |  |
|  |  |

# References

1. <https://www.ccsa.org/what-we-do/student-success>
2. <https://docs.aws.amazon.com/>
3. <https://www.zillow.com/>