High Level Design

Big Game Census Data Visualization

(Big Game Census Data Analysis)

Revision number: 1.0

Last date of revision: 10/11/2021

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 10/11/2021 | HLD-V1.0 | First Version of Complete HLD | Sujitha Sunkara |

**Contents**

[Document Version Control 2](#_bookmark0)

[Abstract 4](#_bookmark1)

1. [Introduction 5](#_bookmark2)
   1. [Purpose of the Document 5](#_TOC_250006)
   2. [Objective of HLD 5](#_TOC_250005)
   3. [Scope of HLD 5](#_TOC_250004)
2. [General Description 6](#_bookmark3)
   1. [Product Perspective & Problem Statement 6](#_bookmark4)
   2. [Data Requirements 6](#_TOC_250003)
   3. [Tools Used 7](#_TOC_250002)
   4. [Constraints 7](#_TOC_250001)
3. [Design Details](#_bookmark5) 8
   1. [Process Flow 8](#_TOC_250000)
   2. [Error Handling / Exception Handling](#_bookmark6) 8
4. [Conclusion](#_bookmark7) 09
5. [References](#_bookmark8) 10

# Abstract

The United States Census Bureau (USCB), formally the Bureau of the Census, is a primary agency of the United States Federal Statistical System in charge of providing data about the American people and economy.

The Census Bureau is advancing with the role of data science and data analytics as the census evolves every year, both in terms of population and economics. Data science aids in understanding how the population, economy, income, poverty, and a variety of surveys vary regionally, in a timely way, and so on, and why. It assists the government in addressing the challenges of backward regions and identifying specific information in the region.

Teams from California and Missouri will compete in the 54th tournament to crown a new football champion. Did you know that this game includes players from 33 different states and five foreign countries? The Big Game Census interactive visualization looks more closely at the players' birthplaces and provides access to Census data points for specific locations.

Census data is advancing with the importance of data science and analytics in a world of expanding new technologies and innovation. Data analysis might assist them gain a new perspective on the census bureau. This study reveals how various studies may assist to understand players' birthplaces and personality features, which can lead to new and better judgments for supporting athletes. Various studies, such as exploratory data analysis and descriptive data analysis, were performed on a variety of use cases in order to get key insights from this data, which would be utilized to make decisions.

# 

# Introduction

This document will be used for documenting High-level designs of project.

## Purpose of the Document

The purpose of this plan is to

* + 1. Describe different design approaches.
    2. Describe different analysis approaches based on variety of Use Cases.
    3. Describe third party components/tools required for the system.
    4. Present complete Process Flow followed for this project.

## Objective of HLD

1. To provide an overview of the entire system.
2. To provide introduction of Problem Perspective & Statement, Data Requirements, Tools used and many more.
3. To provide a module-wise breakup of the entire system.

## Scope of HLD

This HLD covers all areas of system.



# General Description

## Product Perspective & Problem Statement

Many individuals like watching games, and everyone has a favorite. If a new spectator is watching a game and is unsure who to support, they will go through this study of players and their birthplaces, colleges, and so on. In this project, we are examining numerous features with diverse use cases that span many areas of Big Game Census. This not only helps us grasp the relevant links between qualities, but it also allows us to do our own research and come up with our results.

The project's purpose is to do exploratory data analysis, data preprocessing, data cleaning, and imputation, followed using various data visualization techniques to derive meaningful insight from the supplied data. This project seeks to use some excellent Tableau visualizations to improve our visual knowledge of the data.

## Data Requirements

Data Requirement completely depend on our problem.

* + - In this project, to perform analysis, we are using datasets that are provided by iNeuron Company.
    - We make a use of those different datasets as per the requirement and the problem statement.
    - The features which are taken into consideration are:
    - Some of the important features are:

|  |  |
| --- | --- |
| **Name** | **Description** |
| **Player\_BirthPLace** | Birthplace of player |
| **Player\_name** | Name of the player |
| **Player\_team** | Player from which team |
| **Player\_state** | State of the player |
| **Population\_estimates** | Population estimation of player places |

## Tools Used

* + - Tableau desktop
    - Data is connected through Tableau connect
    - Exploratory data analysis is automated by data prep.
    - For visualization of the plots, Tableau sheets are used
    - Dashboards are created using Tableau dashboards
    - Story was created for visualizations
    - GitHub is used as version control system

Chart, scatter chart

Description automatically generated

## Constraints

## The analysis must be user pleasant, the visuals must be tidy and clear, and EDA should be automated as much as feasible to save time. Furthermore, users should not be required to have any coding skills because the insights they seek are described in full with accompanying graphics.

# Design Details

## Process Flow

**Define the Use Cases**

**Import**

**the data**

**Perform**

**Exploratory Data Analysis**

**Data pre-**

**Processing, Data Cleaning & Imputation**

**Visualize and**

**share meaningful insights**

**Analyze the**

**data**

**Handling the**

**Categorical & Numerical Variables**

## Error Handling / Exception Handling:

We created this project in such a manner that the entire script is tested and run numerous times to ensure that no errors arise during the process flow.

In order to reduce misunderstanding, we have also removed any unnecessary warnings.

# Conclusion

In this analysis project, several use cases for the supplied dataset are studied to assist make better judgments and assess player information, as well as their locations and percentage of population change over time. It has been discovered that about 15 players were born in California, with Texas coming in second. We may also get the population of the state as well as information about it with a single click. Tom Brandy of the New England team has played more games than the others. More players from Florida State and West Virginia University are on the Philadelphia club. Michigan and Rutgers have more players than the New England team, while Philadelphia team has more players from the California with an average age of 25, New England team has more players from Texas with an average age if 26. From 2010 to 2017, Texas saw the most population growth. In the United States, there has been little change in population from 2016 and 2017.

# References

[BIG Game Census Data Visualization](https://www.census.gov/library/visualizations/interactive/big-game-census-2020.html)