

Low Level Design Document

Big Game Census Data Visualization

-Aishwarya Srivastava

Written By	Author 1- Aishwarya Srivastava
Document Version	0.1
Last Revised Date	

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	7- Sep - 2021	Author 1	

Reviews:

VERSION	DATE	REVIEWER	COMMENTS
0.1	8-9-2021	Author 1	Unit test cases to be added

Contents

1. Introduction.....	04
1.1 What is Low-Level Design Document?.....	04
1.2 Scope.....	04
2. Architecture.....	05
2.1 Tableau Server Architecture.....	05
2.2 Components of Tableau Server Architecture.....	06
2.3 Summary.....	07
3. Architecture Description.....	08
3.1 Data Description	08
3.2 Web Scrapping.....	09
3.3 Data Transformation	09
3.4 Data insertion into database.....	09
3.5 Connection with SQL server	10
3.6 Export Data from database	12
3.7 Deployment	12
4. Unit test cases	14

1. INTRODUCTION

1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

The Big Game Census dataset is visualized on Tableau Desktop and published on Tableau Public. The dashboard given below can be accessed using the link:

<https://public.tableau.com/app/profile/aishwarya6335/viz/BigGameCensusDashboard/Dashboard1>

2. ARCHITECTURE

2.1 Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

The various layers used in the Tableau server are given in the following architecture diagram.

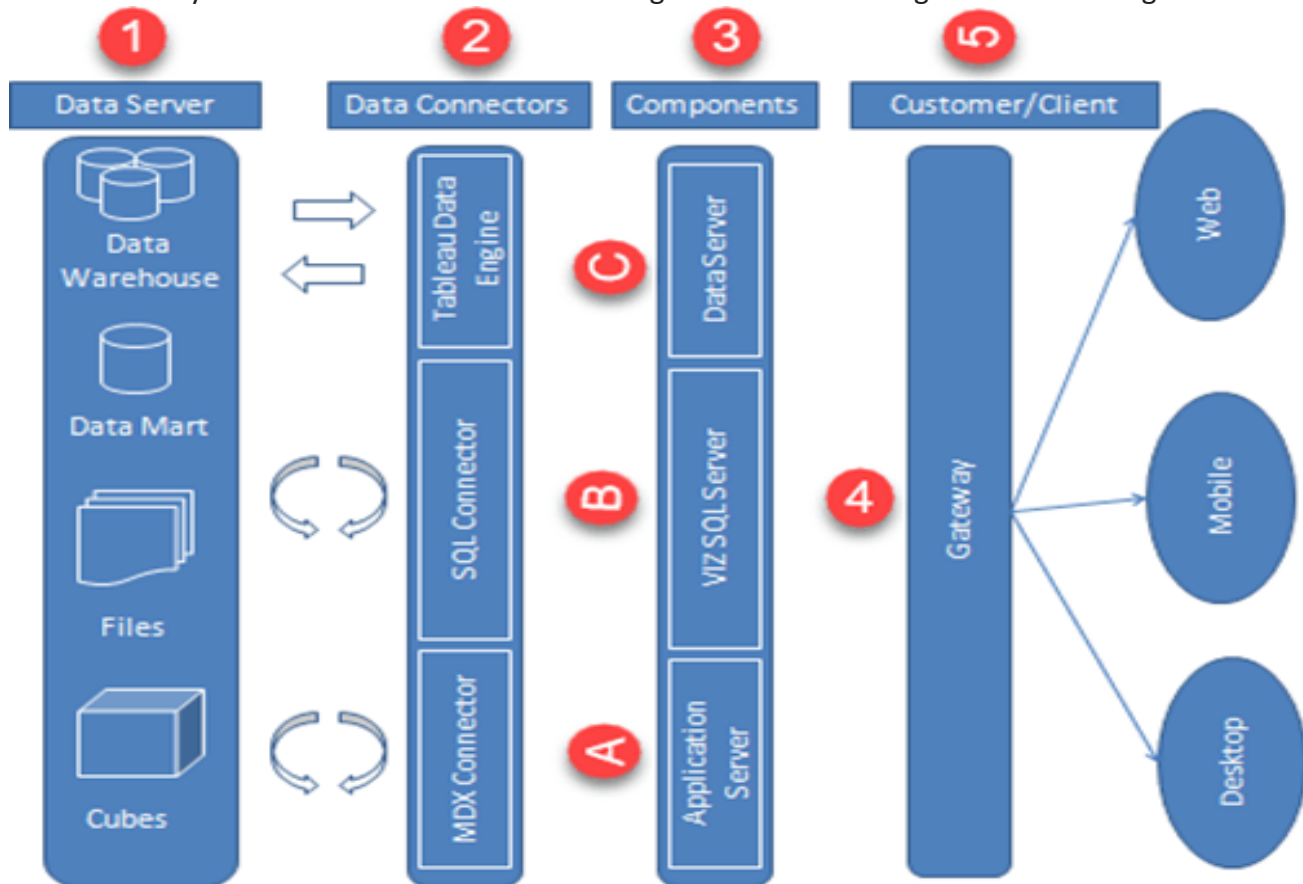


Tableau Architecture Diagram

Let's study the different components of Tableau Architecture

➤ Data Server

The primary component of Tableau Architecture is the Data sources it can connect to it.

Tableau can connect to multiple data sources. These data sources can be on-premise or remotely located. It can connect to a database, excel file, and a web application all at the same time. Tableau can connect data from heterogeneous environments. It can blend the data from multiple data sources. It can also make the relationship between various types of data sources.

➤ Data Connectors

The Data Connectors provide an interface to connect external data sources to Tableau Data Server.

Tableau has in-built ODBC/SQL connector. This ODBC Connector can connect to any databases without using their native connector. Tableau has an option to select both live and extract data. Based on the usage, one can be easily switched between extracted and live data.

- **Live Connection or Real time data:** Tableau can connect to real time data by linking to the external database directly. It uses the infrastructure of existing database system by sending dynamic MDX (Multidimensional Expressions) and SQL statements. This feature can link to the live data with Tableau rather than importing the data. It makes good the investment done by an organization on a fast and optimized database system. In many enterprises, the size of the database is huge and is updated periodically. In those cases, Tableau works as a front-end visualization tool by connecting to the live data.
- **Extracted or In-memory data:** Tableau has an option to extract the data from external data sources. We can make a local copy in the form of tableau extract file. It can extract millions of records in Tableau data engine with a single click. Tableau's data engine uses storage such as RAM, ROM and cache memory to store and process data. Using filters, Tableau can extract few records from a huge dataset. This improves the performance, especially while working on massive datasets. Extracted or in-memory data allows the users to visualize the data offline, without connecting to the data source.

2.2 COMPONENTS OF TABLEAU SERVER

The different components present in a Tableau server are:

- Application Server
- VizQL Server
- Data Server

A) Application Server:

The application server is used to provide the authentications and authorizations. It handles the administration and permission for web and mobile interfaces. It assures security by recording each session id on Tableau Server. The administrator can configure the default timeout of the session in the server.

B) VizQL Server:

VizQL server is used to convert the queries from the data source into visualizations. Once the client request is forwarded to VizQL process, it sends the query directly to data source and retrieves information in the form of images. This image or visualization is presented to the user. Tableau server creates a cache of visualization to reduce the load time. The cache can be shared across many users who have the permission to view the visualization.

C) Data Server:

Data server is used to manage and store the data from external data sources. It is a central data management system. It provides metadata management, data security, data storage, data connection and driver requirements. It stores the relevant details of data set such as metadata, calculated fields, sets, groups, and parameters. The data source could extract data as well make live connections to external data sources.

➤ Gateway

The gateway channelizes the requests from users to Tableau components. When the client makes a request, it is forwarded to external load balancer for processing. The gateway works as a distributor of processes to various components. In case of absence of external load balancer, gateway also works as a load balancer. For single server configuration, one primary server or gateway manages all the processes. For multiple server configurations, one physical system works as primary server while others are used as worker servers. Only one machine can be used as a primary server in Tableau Server environment.

➤ Clients

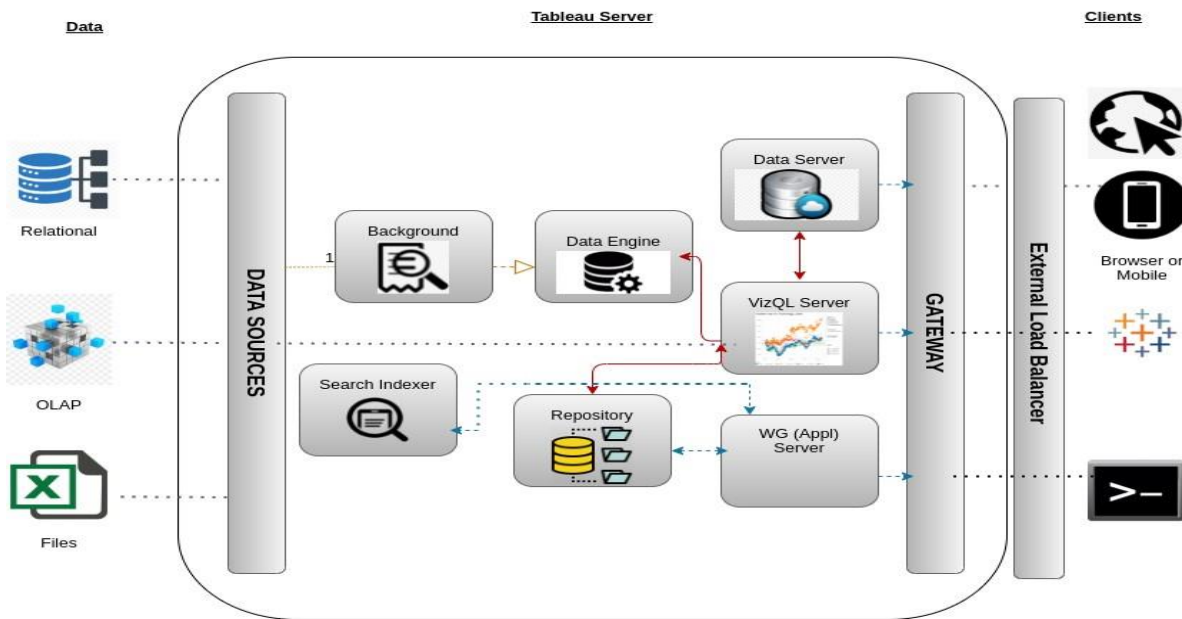
The dashboards and visualizations in Tableau server can be viewed and edited using different clients. The Clients are Tableau Desktop, web browser and mobile applications.

Clients	Environment
Tableau Desktop	Tableau Desktop is a business analytics tool. It helps to create, view and publish dashboards in Tableau Server. Users can access various data sources and build visualizations in Tableau Desktop.
Mobile	The dashboards from the server can be interactively visualized using mobile browsers and applications. The browser and application can be used to view and edit the contents in the workbook.
Web	Web browsers such as Google Chrome, Safari, Firefox and internet explorer support the Tableau server. The contents and visualizations in the dashboard can be edited through these web browsers.

2.3 Summary

- The architecture of Tableau Server is designed to connect different data sources securely.
- Data Server is the first layer in the architecture. It helps Tableau to connect data in various heterogeneous environments.
- Data connector is the second layer. It helps to connect to various databases using its ODBC connector.
- Tableau can connect the real time live data by connecting the database directly. It can also extract a local copy of data through its in-memory data store for faster processing.
- The components such as Application Server, VizQL Server, and data server act as the third layer.
- The application server is used for authentications and authorizations.
- VizQL is used to convert the SQL query into visualizations.
- Data Server is a centralized data management system used in the architecture.
- A gateway is used to distribute the processes into different components.
- The fourth layer of the architecture are the clients such as Tableau Desktop, web and Mobile.
- Tableau Server is internally managed by the multiple server processes.
- The following diagram shows Tableau Server's architecture:

Tableau Communication Flow



3. Architecture Description

3.1. Data Description

The Dataset of Big Game Census data visualization contains the data about the game, teams, players, their birthplaces and population of their birthplaces, etc. Dataset contains the following columns:

1. Player Name- contains the name of the player
2. Player Jersey Number- contains the jersey number of each player
3. Player Position- Player at which the player plays in the team
4. Player Age- Age of each player
5. Player Weight (lbs.)- Weight of each player in pounds
6. Years Played- Number of years the player has played
7. Player Birthplace (city,town,etc.)- Birthplace, including city and town, of each player.
8. Player Birth State- State of birth of each player
9. Player Birthplace (Combo)- City and State of birth of each player
10. Player College-Name of the college the player went to
11. Player Team-Name of the team the player plays for
12. Conference- Has 2 values, AFC and NFC

13. 2016 Population Estimates (except where otherwise noted)- Estimated population of all birthplaces for the year 2016
14. State GEO ID- Last 2 digits of the corresponding full GEO ID is the state GEO ID
15. Full GEO ID- GEO ID of all the birthplaces
16. Latitude (player birthplace)- contains the latitude of the birthplace of the player
17. Longitude (player birthplace)- contains the longitude of the birthplace of the player
18. Number from city- Contains the number of entries from all cities
19. Number of Records-contains the number of records
20. American Factfinder link for more census data points-contains the link to more census data point
21. Quickfacts Link-contains the quick facts link for all cities
22. State Data Link-contains the state data link of all states
23. Source (Population States 2017)-contains the name of the source from where the population is extracted for all states for the year 2017
24. Birthplace (Population Data Source)- contains the name of the source from where the population is extracted for all birthplaces

3.2. Web Scrapping

Web scrapping is a technique to automatically extract content and data from websites using bots. It is also known as web data extraction or web harvesting. Web scrapping is made simple now days, many tools are used for web scrapping. Some of python libraries used for web scrapping are Beautiful Soup, Scrapy, Selenium, etc.

3.3. Data Transformation

In the Transformation Process, we will convert our original datasets with other necessary attributes format. And will merge it with the Scrapped dataset.

3.4. Data Insertion into Database

- a. Database Creation and connection - Create a database with name passed. If the database is already created, open the connection to the database.
- b. Table creation in the database.
- c. Insertion of files in the table

3.5. Make the SQL connection and set up the data source

Step 1: Configuring Tableau

Launch Tableau on your workstation and select SQL Server from the connect column on the left. This will open a dialogue box where you need to provide the connection details for SQL Server.

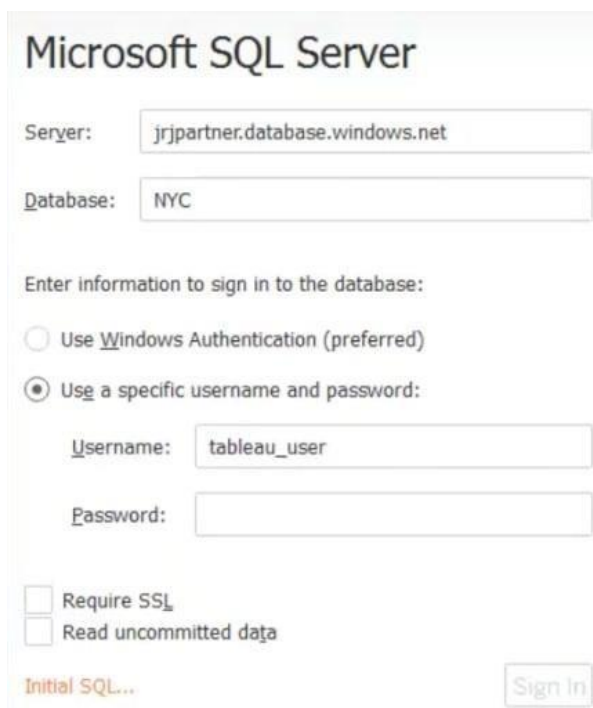
To connect with tableau, you will need to provide information about the server which hosts your database. If you want to connect to a contained database, you can also specify the name of the database.

To connect with a port other than the default port, you need to specify the port and server as follows:

```
<server_name><port_number>
```

Example query: my_server 8051

There are two ways in which you can sign-in to the server, either by using Windows authentication or by using the username and password. Using the username and password becomes a must if you're working with a password-protected server in a non-Kerberos environment.

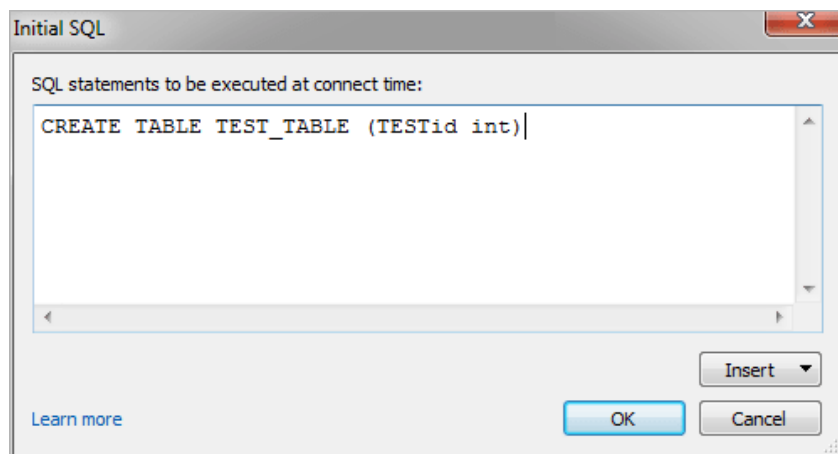


The screenshot shows the 'Microsoft SQL Server' connection dialog box. It has two input fields at the top: 'Server:' with the value 'jrjpartner.database.windows.net' and 'Database:' with the value 'NYC'. Below these is the instruction 'Enter information to sign in to the database:'. There are two radio button options: 'Use Windows Authentication (preferred)' which is unselected, and 'Use a specific username and password:' which is selected. Under the selected option, there are two input fields: 'Username:' with the value 'tableau_user' and 'Password:' which is empty. At the bottom left, there are two unchecked checkboxes: 'Require SSL' and 'Read uncommitted data'. At the bottom right, there is a 'Sign in' button. In the bottom left corner, there is a link that says 'Initial SQL...'. The entire dialog box has a light gray background and rounded corners.

Click on Sign in to establish a connection. This will enable a connection without SSL. To establish an SSL enabled connection, click the Require SSL checkbox before you sign in.

SQL Server provides an option to let the user queries access the modified rows even before they have been committed. This option is called Read Uncommitted data. It saves time by preventing complex queries such as extract refreshes from locking the database and causing a delay. If this option is unchecked, Tableau makes use of default isolation levels.

If you want to run a specific SQL command every-time a new connection is established, you can use the Initial SQL option. This will open a dialogue box, where you can specify your desired SQL query.

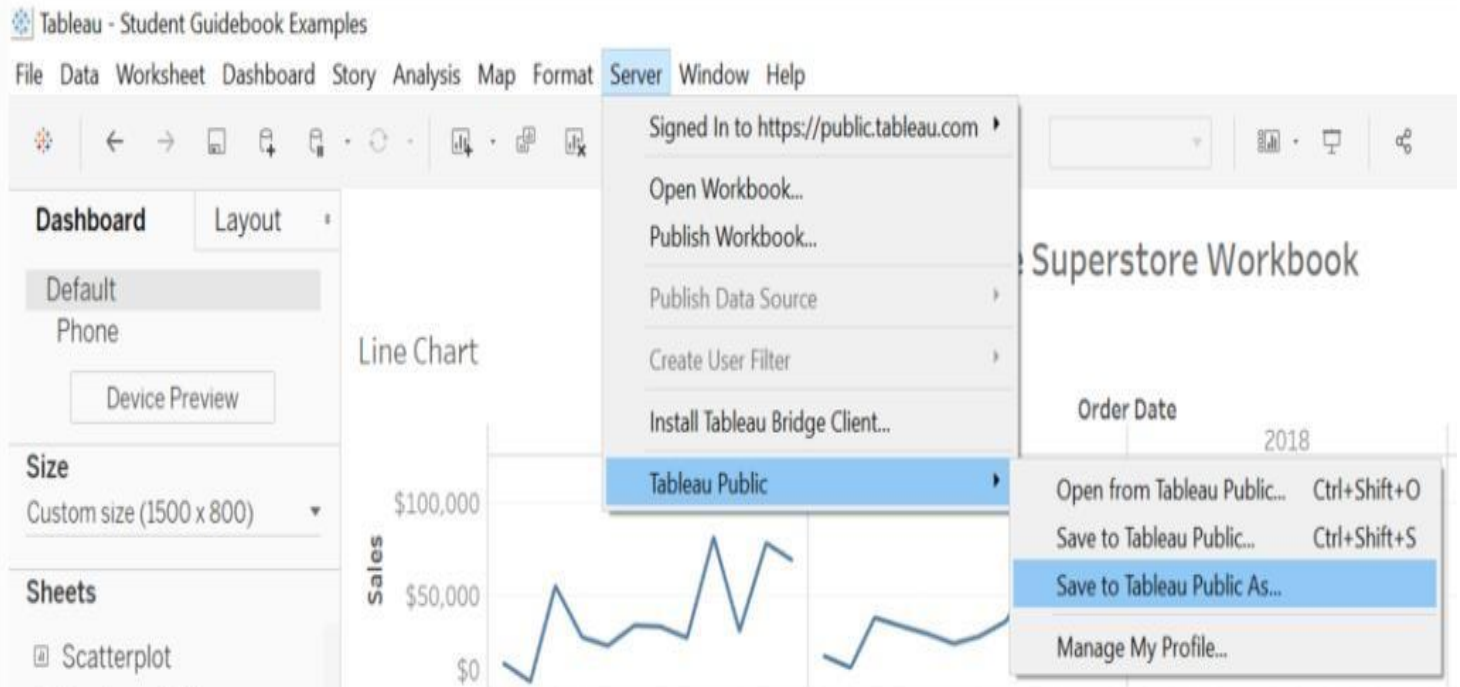


Step 2: Configuring Data Source

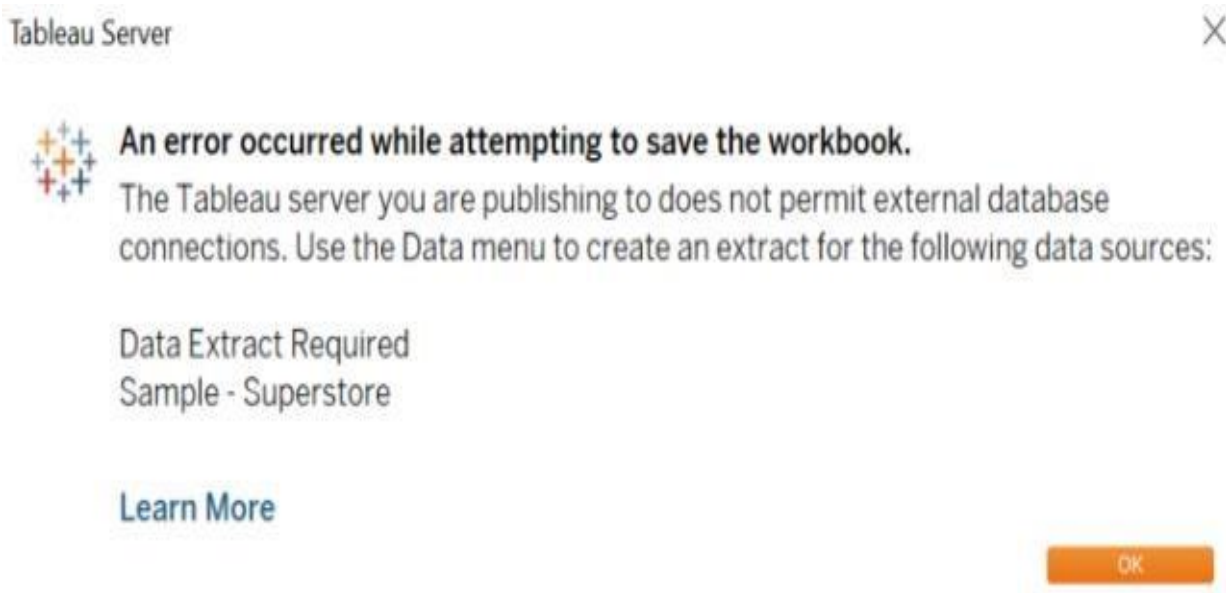
The data source page loads up after configuring the Tableau connector and successfully signing in. This is how the page looks like:



Select the data source name option and give a unique name to the database you are using. It's considered a good practice to have a unique name as it makes it much easier for users to identify the database from which data is being fetched.



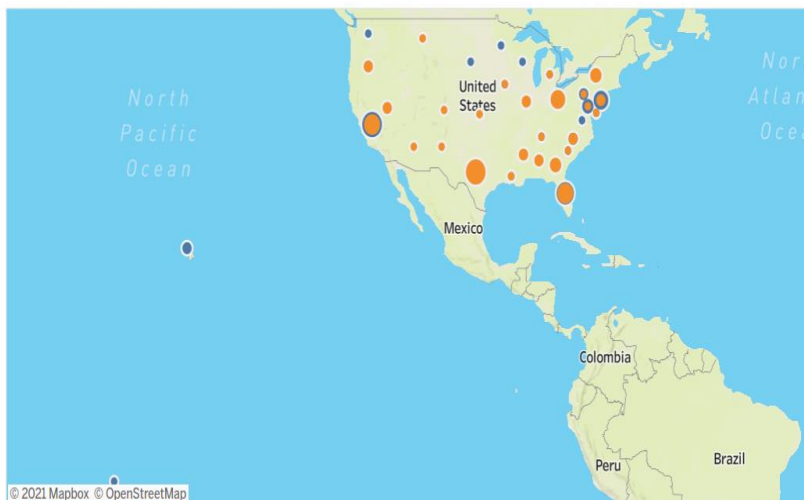
Next, fill out the title you want your viz to have and click “save”.



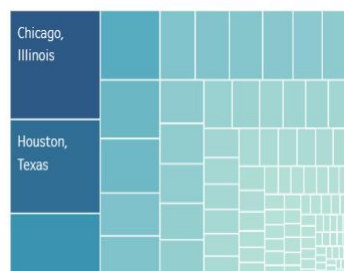
This message means that your connection to the Sample-Superstore data set is a live connection. Tableau Public cannot host live connections, so you'll need to convert your connection to an extract (like a frozen screenshot of your data).

Here in the below screenshot, we can see that our workbook has been published to tableau public.

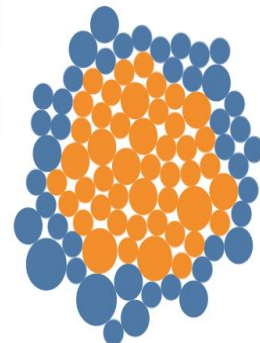
Is someone from your hometown in the game?



Player Birthplace with its 2016 Population Estimates



All Colleges with No. of Players playing for different Teams



2016 Population Estima...



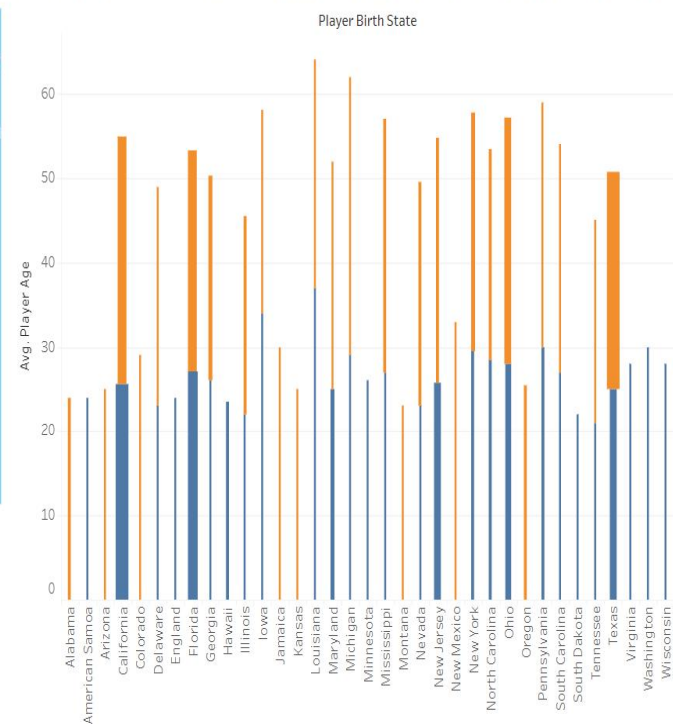
Player Team



Conference



All States with Number of Players, Players Teams & Average Player Age



All Teams with their Avg. Player Age, Avg. Years Played, Median Player Weight & Conference

Player Team	Avg. Player Age	Avg. Years Played	Median Player Weight (lbs.)
New	27.0	5.7	237.0
Philadelphia	26.5	5.7	238.5

Go to Settings to activate Windows.

3. Unit Test Cases

TEST CASE DESCRIPTION	EXPECTED RESULTS
Is someone from your hometown in the game?	When hovered over the map, one can see the number of players from a particular state along with the team to which they belong.
Average Player Age vs Player Birth State	When hovered over the visualization, all States with number of players, player teams & average player age is shown.
Player Team vs Player College	A bubble chart is used here which shows all the colleges with number of players from different teams.
All teams information	A table is used to here which displays the player teams, avg. player age, avg. years played, median player weight and conference to which the players belong.
Player birthplace with population estimates	A treemap shows all the birthplaces having big population estimates with big and dark coloured squares and states having small population estimates are shown with small and light coloured squares.