**Business Intelligence Lab**

**Experiment 10**

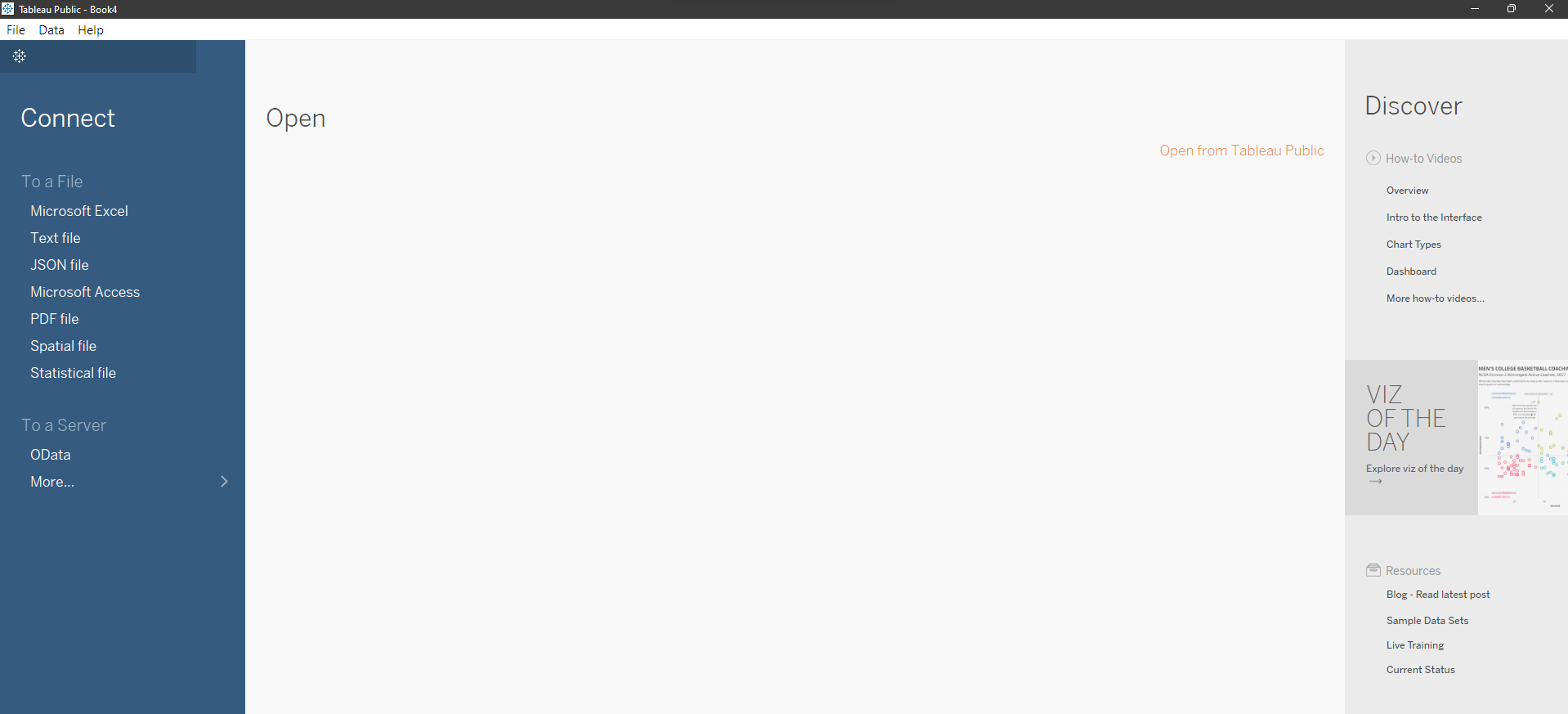
**Aim:** Experiment to study any one BI tool such as Pentaho, Tableau and QlikView

**Theory**:

Comparing the following Business Intelligence tools: Pentaho, Tableau and QlikView:

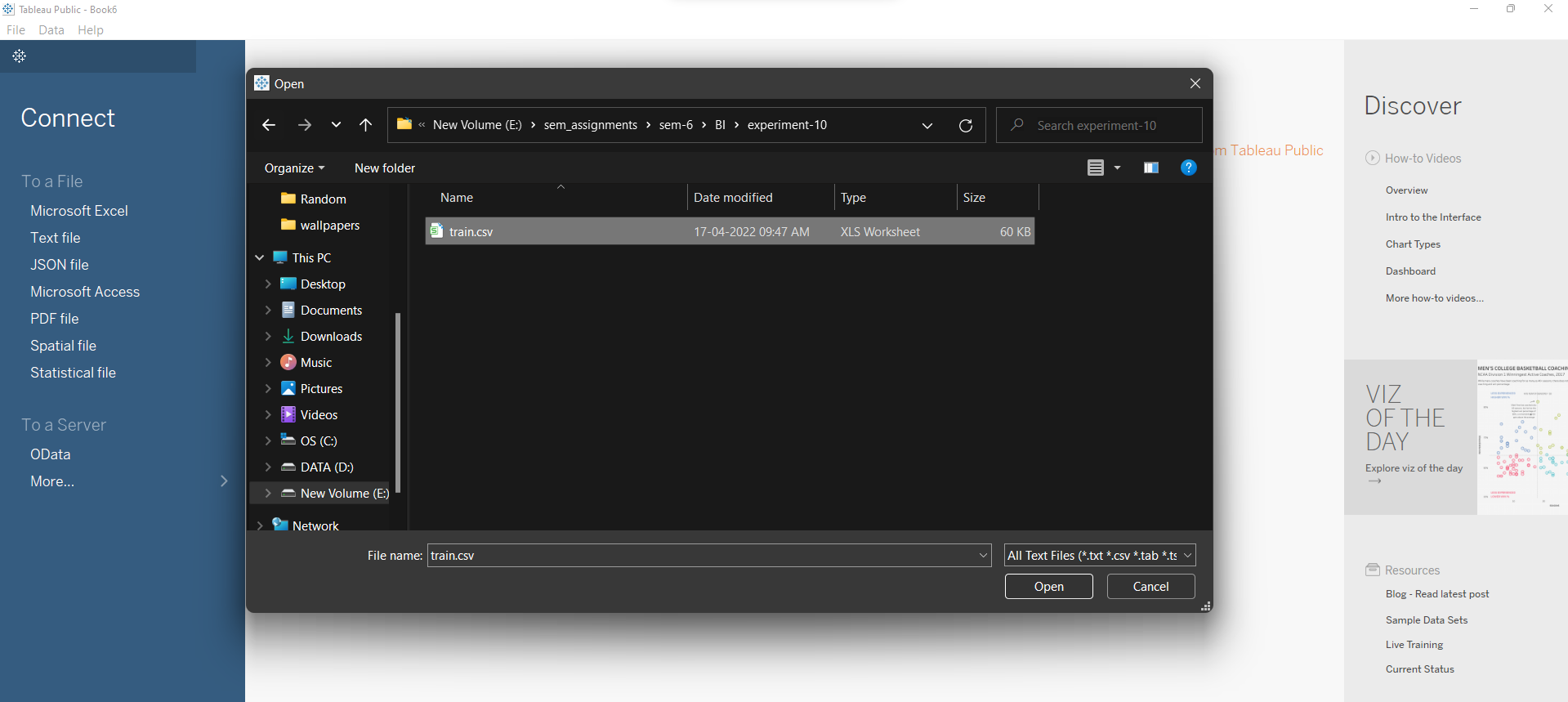
| **Features** | **Pentaho** | **Tableau** | **QlikView** |
| --- | --- | --- | --- |
| **Dashboards and Data Visualization** | Pentaho’s Dashboard Designer helps create interactive visualizations and dashboards with a 360-degree view of data through dynamic filter controls and content linking. | Tableau’s Dashboard Starters make life easier for you by automatically creating dashboards after connecting to popular sources. Can build your own visualizations or utilize those made by other users. | It shows your key metrics’ performance with various visualizations and charts. Its video player visualization lets you embed YouTube videos directly into Qlik Sense apps. |
| **Reporting** | Pentaho’s Report Designer helps create print-ready, interactive reports through easy drag and drop, font selection, and formatting and resizing elements. Users can share reports in HTML, PDF, Excel or CSV format through email. | Its Ask Data module facilitates faster analysis and reporting with natural language querying. Versioning is available, so you can see what has changed from last time. You can subscribe to get reports via email in image or PDF format. | Automatic report delivery isn’t built-in; you need an additional module, Qlik NPrinting, to do the job. Or, you can access reports directly from the Qlik Sense Hub. The tool supports versioning but with third-party integrations. |
| **Data Source Connectivity** | Pentaho Data Integration (PDI), formerly known as KETTLE, connects to more than 40 data sources. It can import data from any file type but connects to only two SaaS platforms – Google Analytics and Salesforce. Its Data Source Wizard allows users to add custom data sources and define them for the Pentaho Server. | It has built-in connectors for Amazon Redshift, Cloudera, Google Analytics, Microsoft Excel, MySQL and more, or you can create your own. It doesn’t connect to project management, payment processing, enterprise messaging and eCommerce platforms out of the box, only through partner integrations. | Qlik Sense also comes equipped with native connectors. You can download any that aren’t available natively from the Qlik website. The vendor continues to expand support for newer sources, the latest ones being the Databricks ODBC and the Azure Synapse connector. |
| **Embedded Analytics** | Based on an open REST-based API, Pentaho supports multi-tenant deployment for reduced complexity and cost. It offers numerous options to embed real-time reports and tailored dashboards into applications and web pages. | It also supports predictive analytics and can work with MATLAB, R and Python for regression analysis. The tool provides built-in date/time functions for comparisons like year-over-year growth and moving averages. | The tool supports bi-variate linear analysis and forecasting through its R plugin. A third-party extension, the Vizlib Line Chart, helps you predict trends and opportunities from historical and near-real-time information. |
| **Augmented Analytics** | Pentaho does not provide augmented analytics. | Its Insight Advisor provides context-aware suggestions, automation and natural language processing. Ask questions in natural language and view the responses as visualizations with Ask Data. | The Qlik Cognitive Engine speeds up preparation and authoring functions with associative recommendations. Smart Search lets you search across the entire data set out of the box. |
| **IoT Analytics** | Pentaho collects data from log files, social media, financial data, telemetry, IoT systems and more. As part of data integration, it operationalizes Python, R, Scala or Weka machine learning models, including those that use scikit-learn, Tensorflow and Keras libraries. | Usage and performance metrics from internet-connected sensors and devices like heavy machinery help minimize defects and monitor field assets in real time. You can prolong the life of your hardware by preempting breakdowns and staying on schedule with preventive maintenance. | It provides a data replication module to accelerate asset ingestion. Working with the Qlik Data Integration Platform, it pulls assets from databases, mainframes and SAP systems and delivers them to streaming systems, data warehouses and data lakes. |
| **Database Security** | Pentaho provides big data security through Kerberos authentication and secure impersonation for connecting to a server cluster. Editor, they can grant and revoke table, row, column and object-level access to data. | Vendor-built open-source tools, LogShark and TabMon, analyze Tableau Server’s usage and performance. Protect your live assets by hiding published dashboards behind logins. Tableau offers security at the row and object level. | Qlik lets you capture CPU and RAM usage and activity metrics through Telemetry Logging. It prevents injection attacks and Cross-Site Scripting (XSS) through the Content Security Policy (CSP) Level 2. |
| **Native Mobile Apps** | Pentaho does not offer mobile support. | Its mobile app allows you to browse, search for and scroll through dashboards on your mobile devices. You can preview and interact with your visualizations and workbooks when offline. Collaboration isn’t available – you can’t add comments or share the annotated mobile screen with others. | You can access Qlik Sense apps and mashups on mobile with all the features – visualization, creation, analysis, collaboration and administration. Add context to analytics with compelling narratives and create active discussions around business assets through collaboration. |

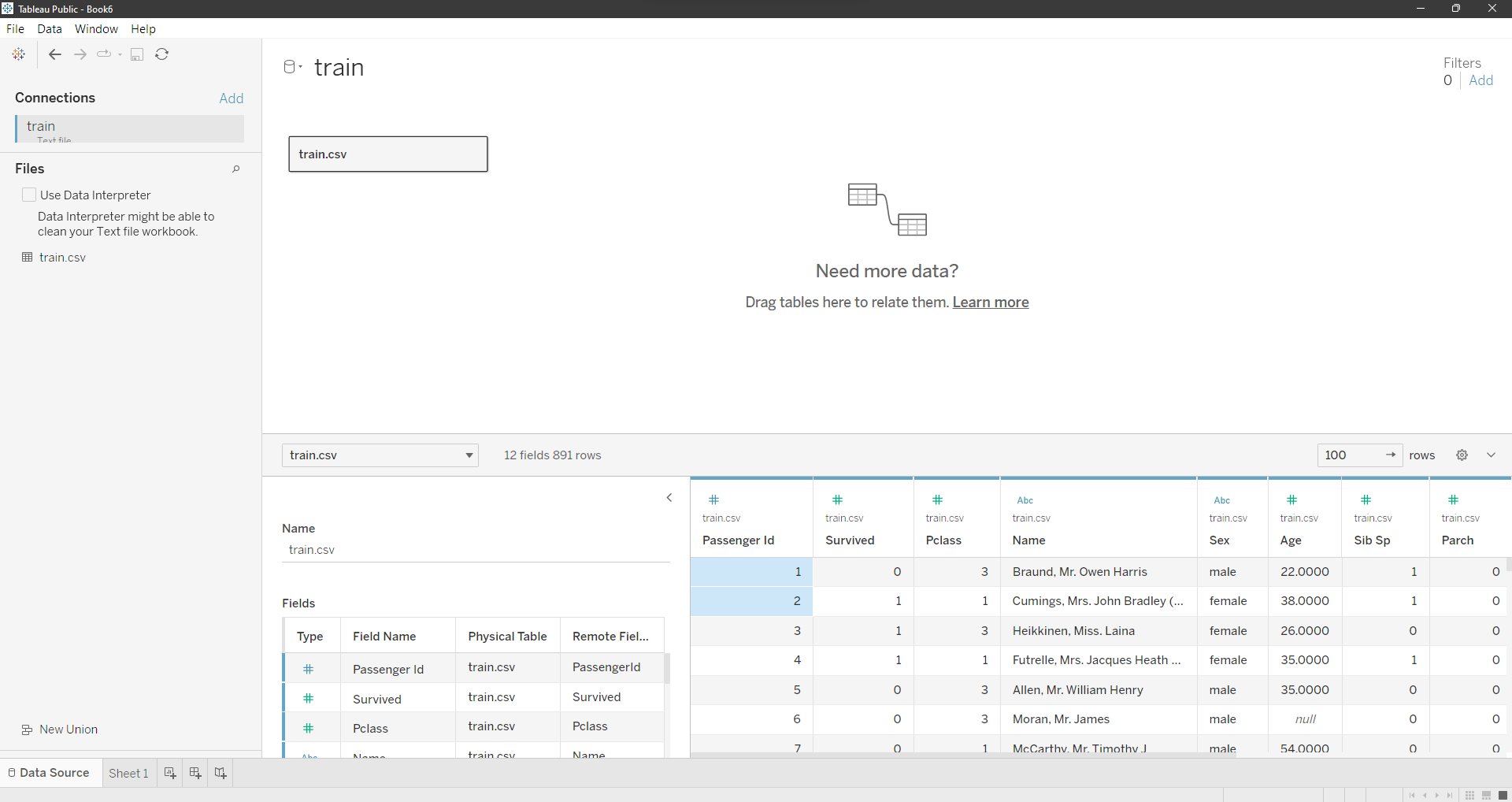
**Output**



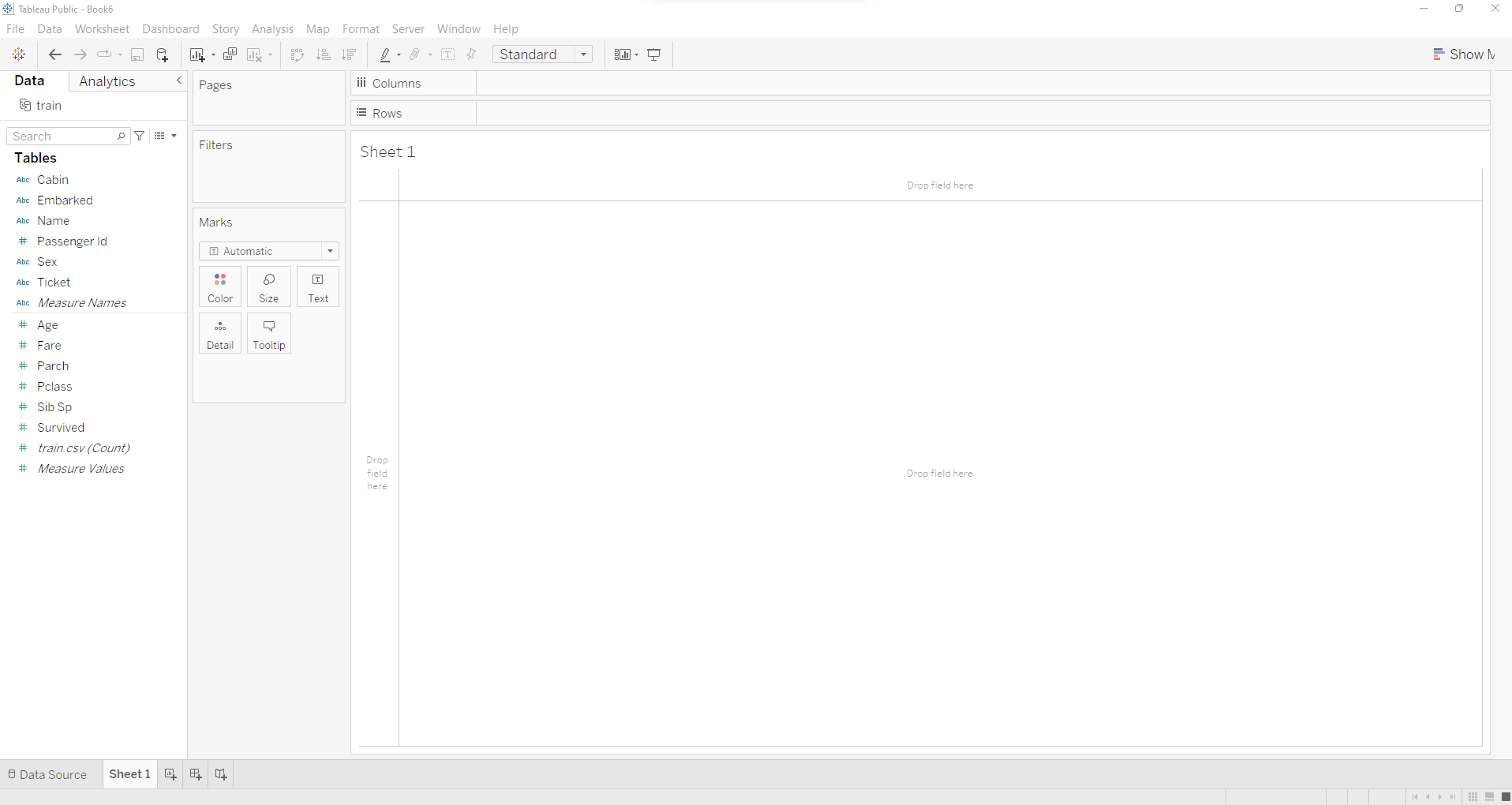
**Load dataset in the tool**

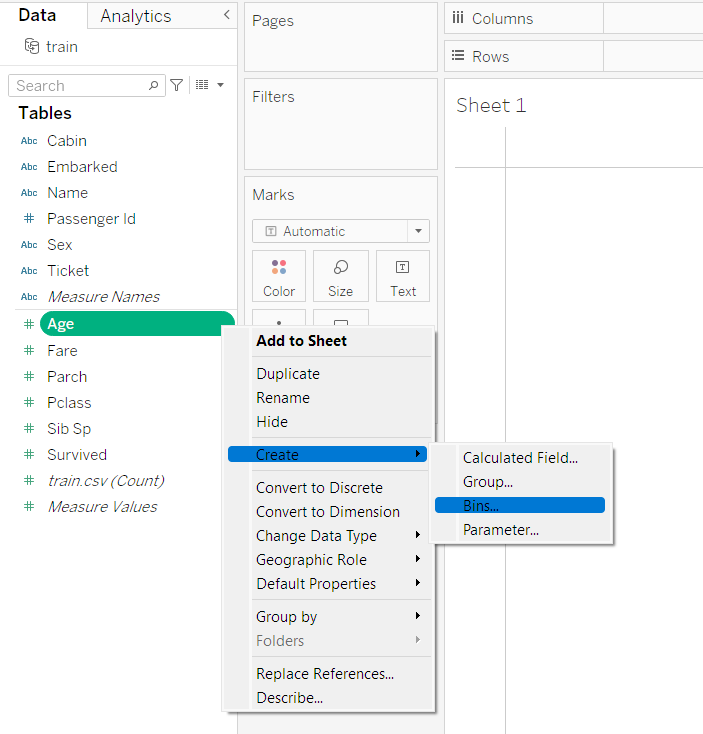
1. Left hand section of Tableau home page, under connect section choose the option suitable to load your data set
2. For loading CSV file, click on “Text file”
3. Select the CSV from windows explorer
4. Dataset is loaded on tableau



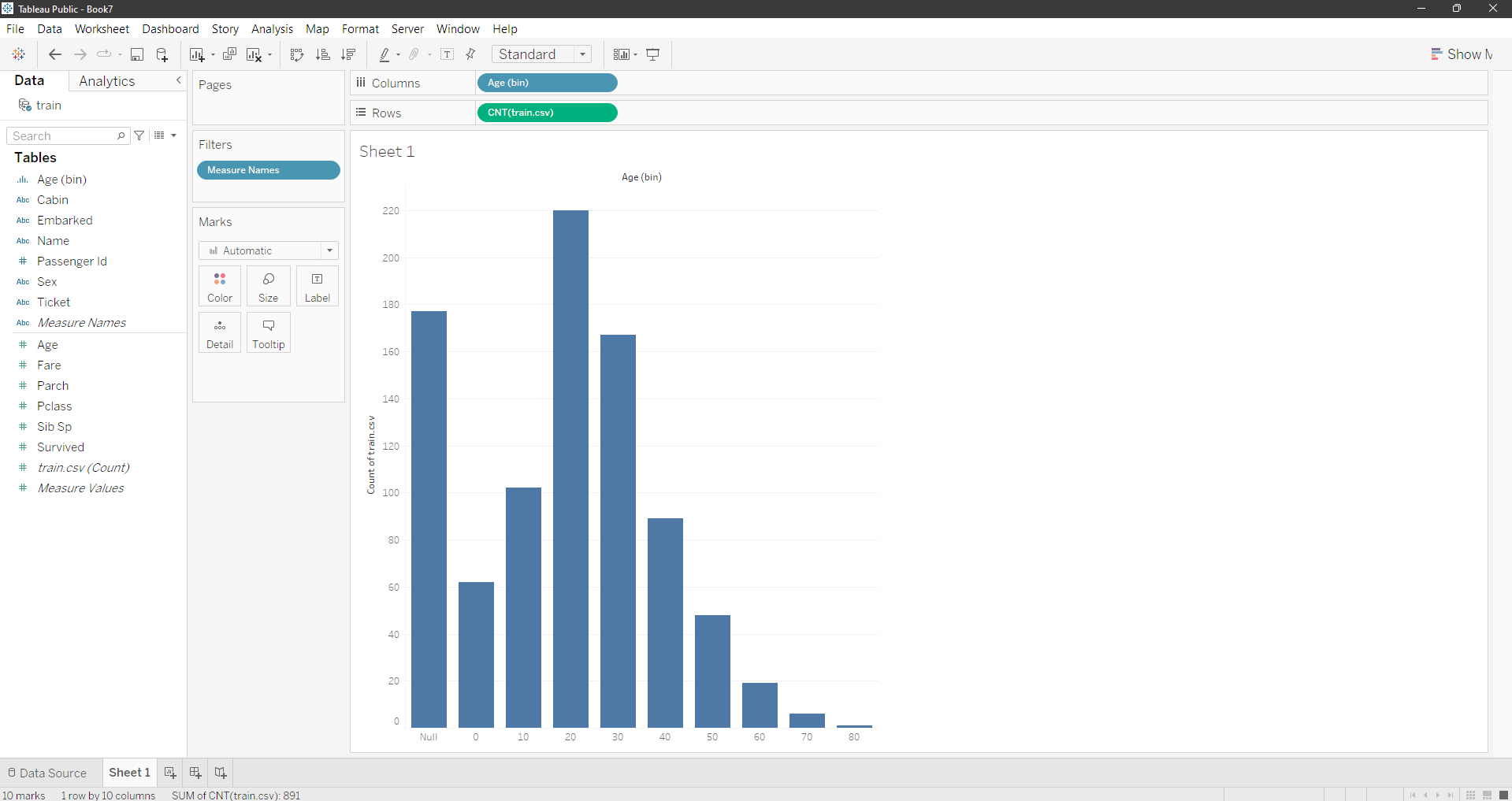


**Visualization**

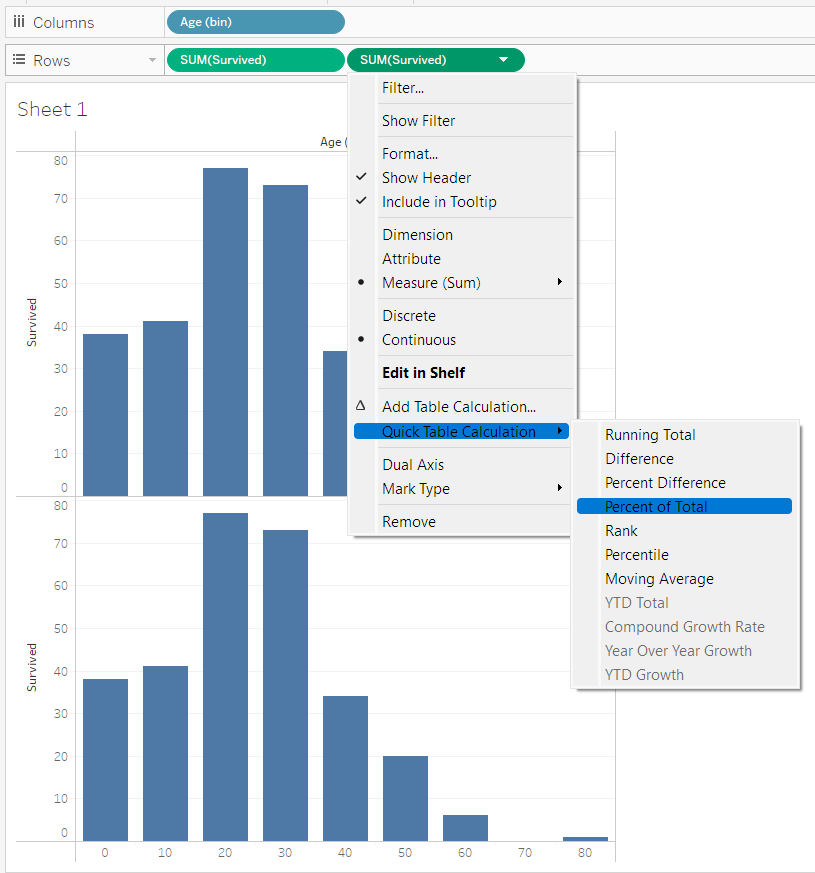
1. Click on Sheet 1 at the bottom to open visualization window  
   
2. On the left hand side we can see Tables section, which contain **Dimension** at top and **Measures** at bottom
   1. Dimensions contain qualitative values (such as names, dates, or geographical data i.e categorical). You can use dimensions to categorize, segment, and reveal the details in your data. Dimensions affect the level of detail in the view
   2. Measures contain numeric, quantitative values that you can measure. Measures can be aggregated. When you drag a measure into the view, Tableau applies an aggregation to that measure (by default)
3. First we will visualize the correlation between the number of people survived and their age, steps to do that are as follows
   1. Create dimension of Age measure by clicking on Age drop down -> create -> bins



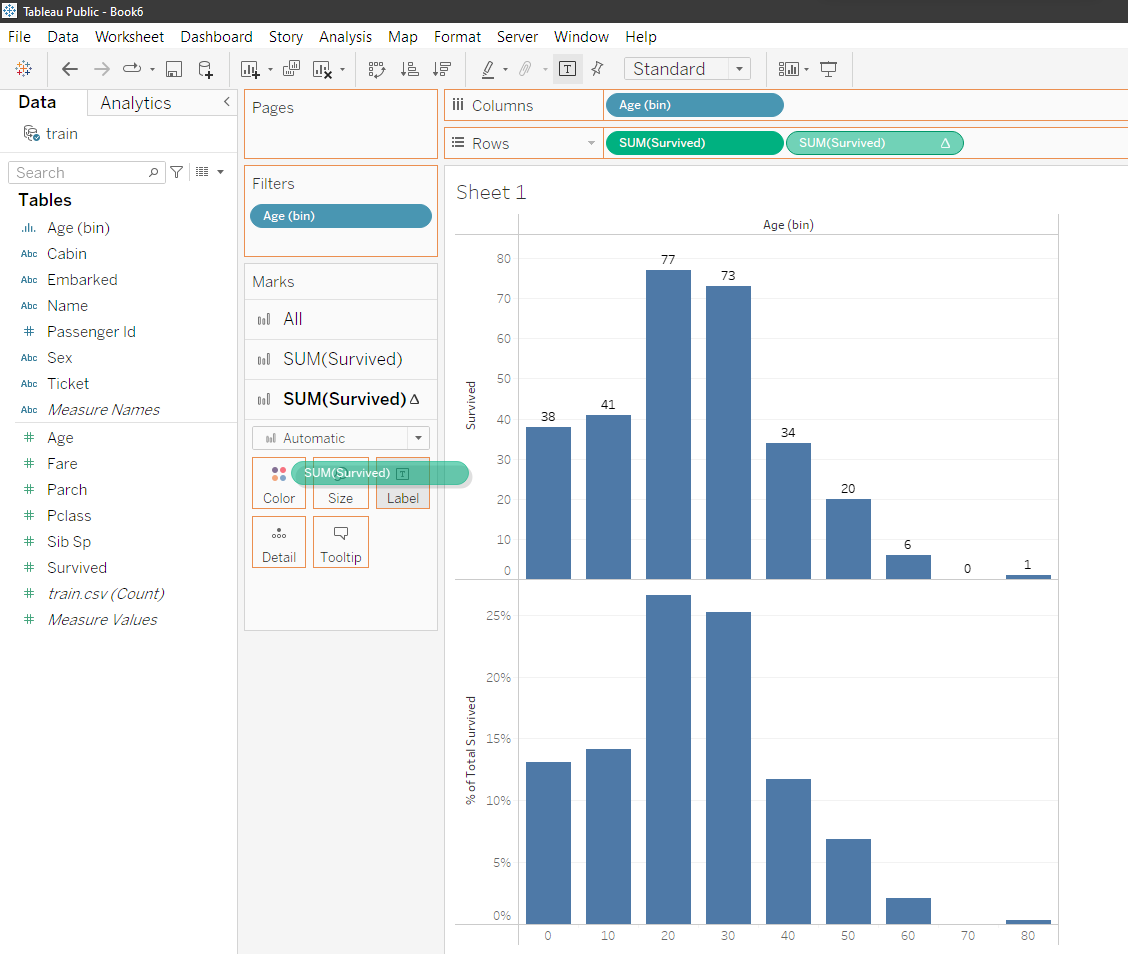
* 1. After the bin is created, we visualize Age against number of records, to do this, click and drag Age (bin) to columns and train.csv (count) to Rows as shown



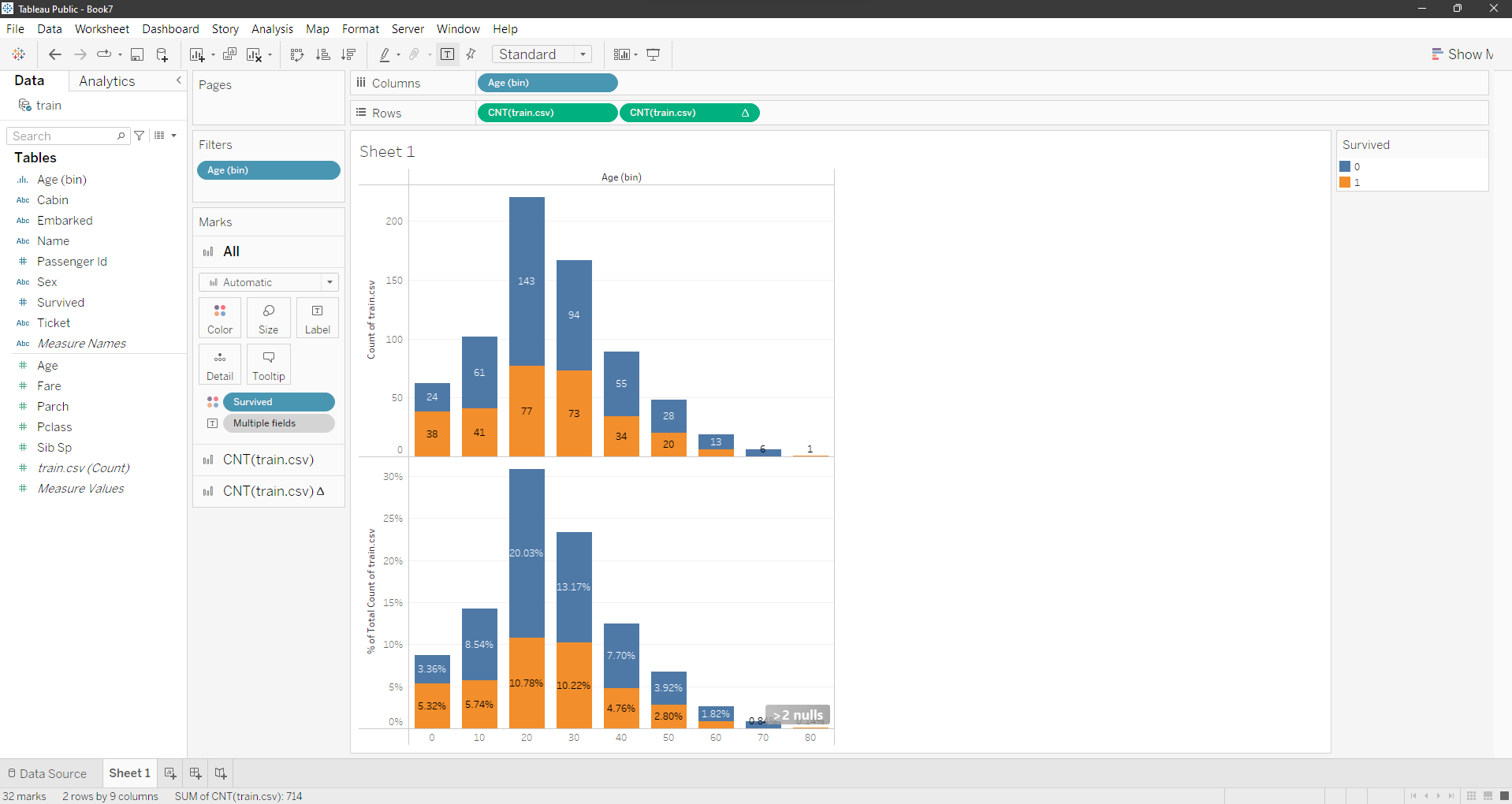
* 1. Here we can see that there are a lot of null values. To remove the null values, ctrl+click on Age (bin) and drag it to the filters section and un-select Null -> apply -> okay. Null values are handled
  2. To add more context to the data, to see the percentage of survivors w.r.t their age, we add “survived” to rows again, another graph is formed. In the newly added row, click on dropdown -> Quick table calculation -> percent total. We can now see the percentage.



* 1. To add labels, ctrl+click and drag the CNT to respective mask in label section as shown

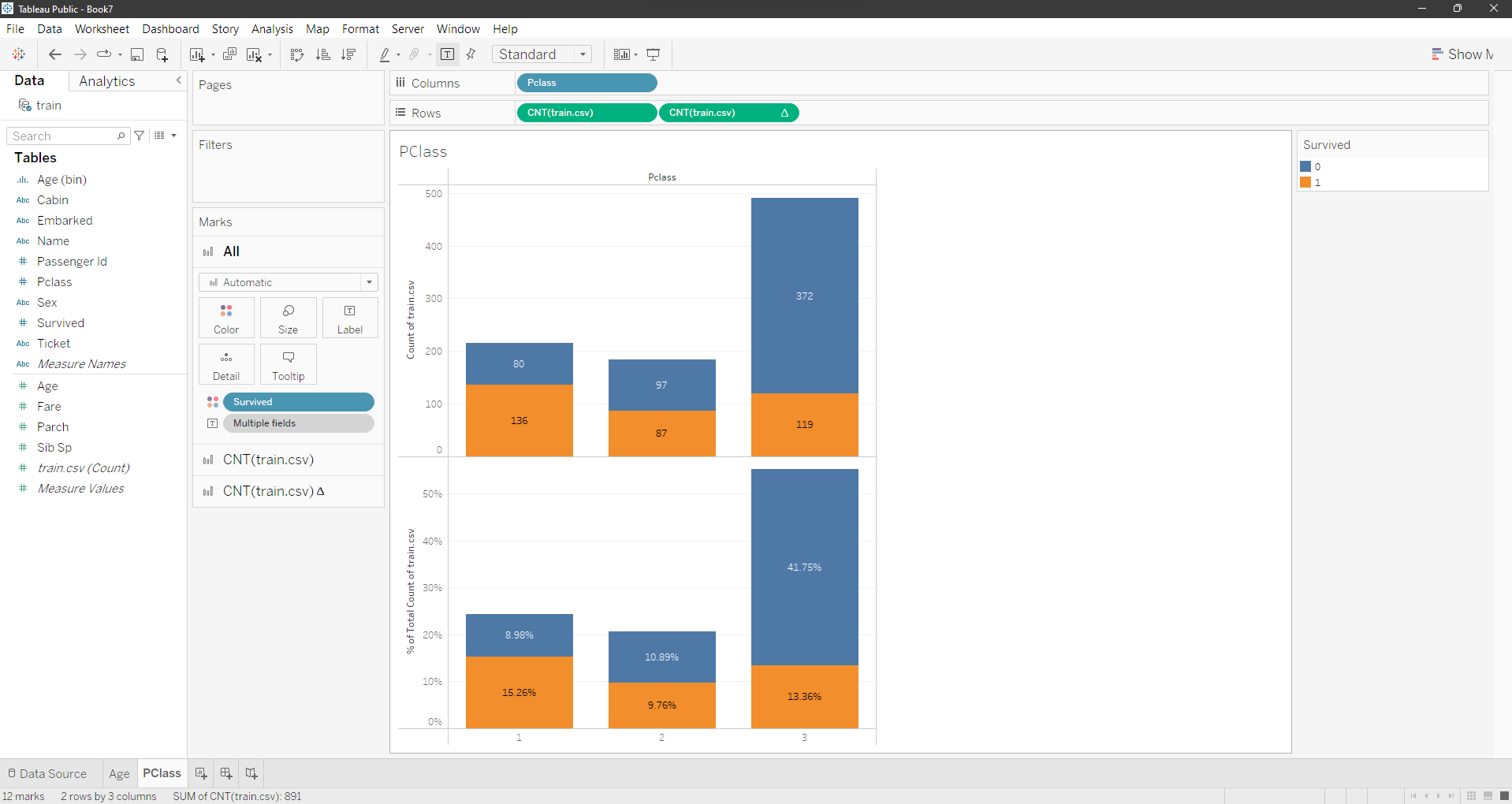


* 1. To further make the visualization better, we convert “Survived” to dimension and then drag it onto Colors section of All in Marks

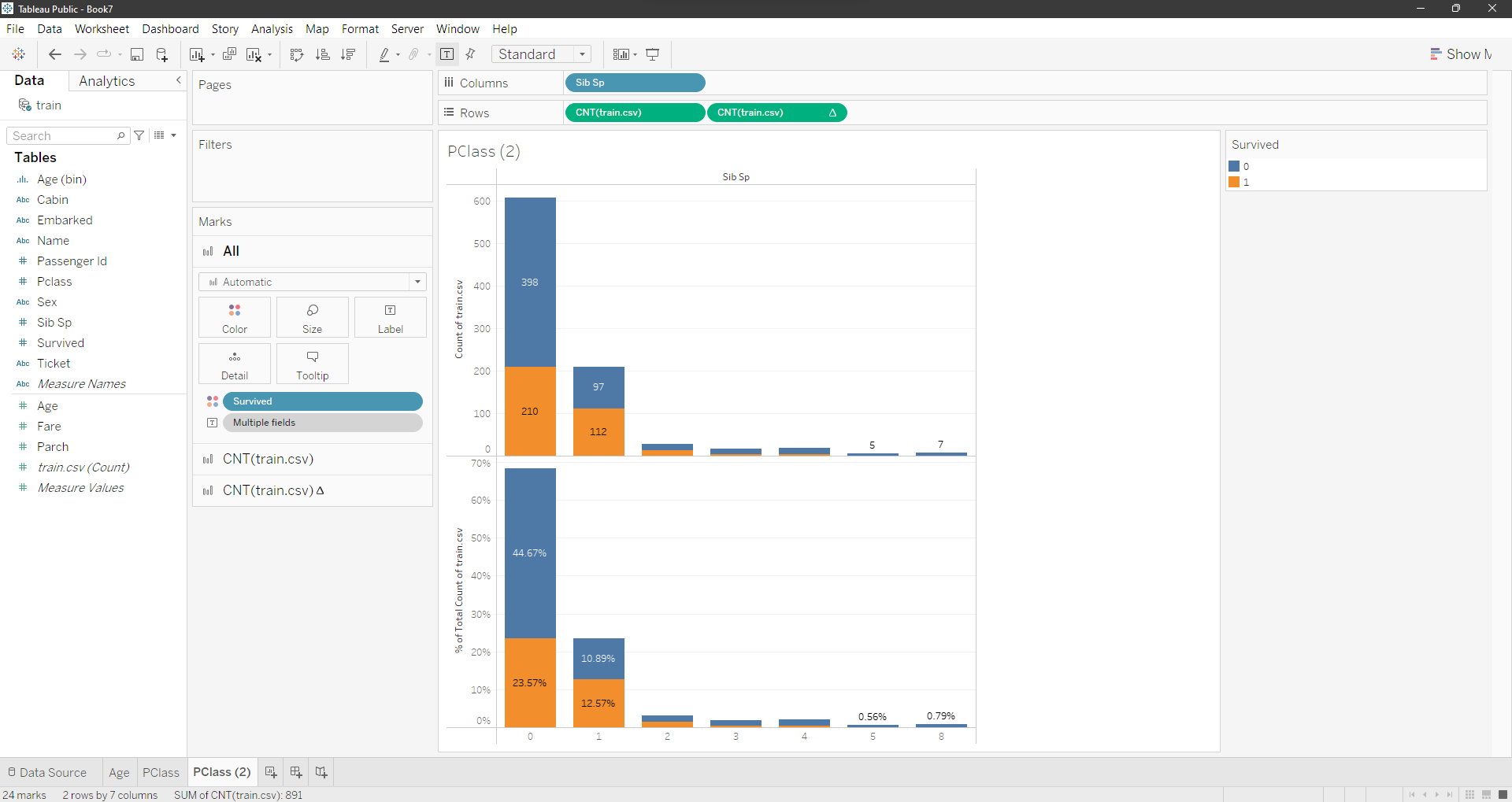


* 1. Here, oranges survive.
  2. Now that we have one of the visualizations ready we can duplicate and create many more.

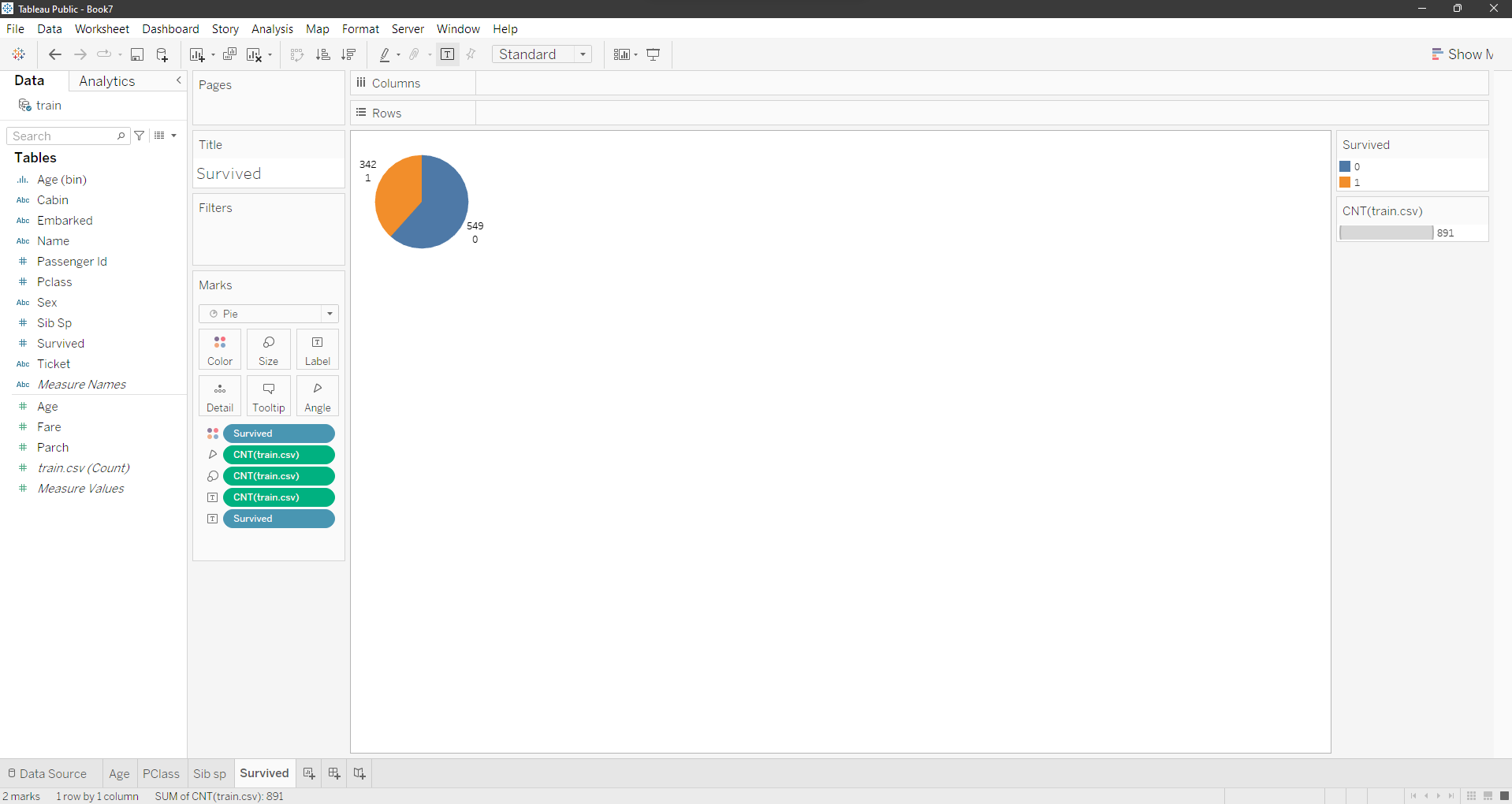
1. After visualizing the number of survivors against age we now visualize it against different factors. Follow the same steps above.
2. Visualizing survivors against passenger class



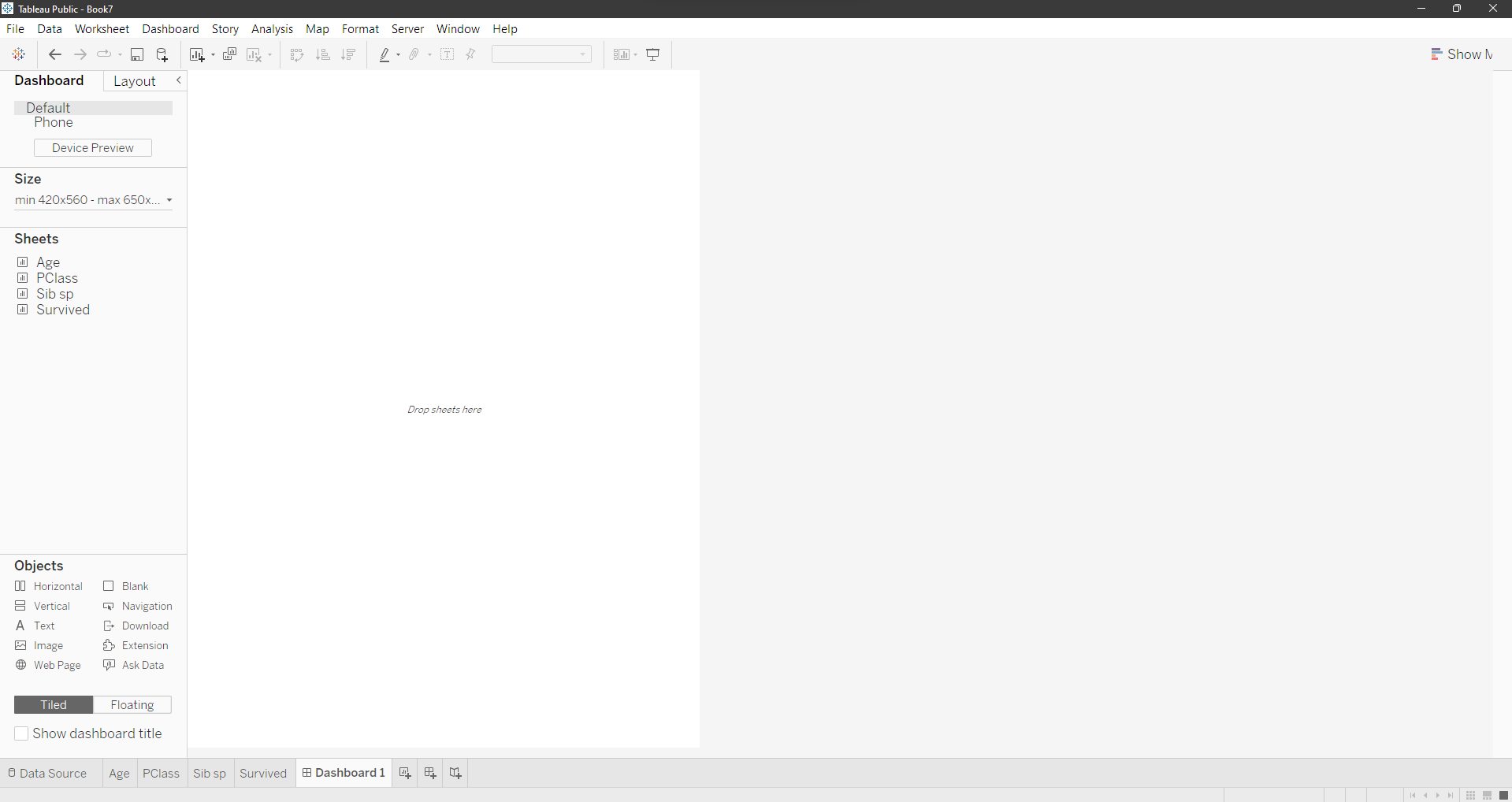
1. Visualizing survivors against Sib Sp class. sibsp: The dataset defines family relations in this way… Sibling = brother, sister, stepbrother, stepsister Spouse = husband, wife (mistresses and fiancés were ignored)



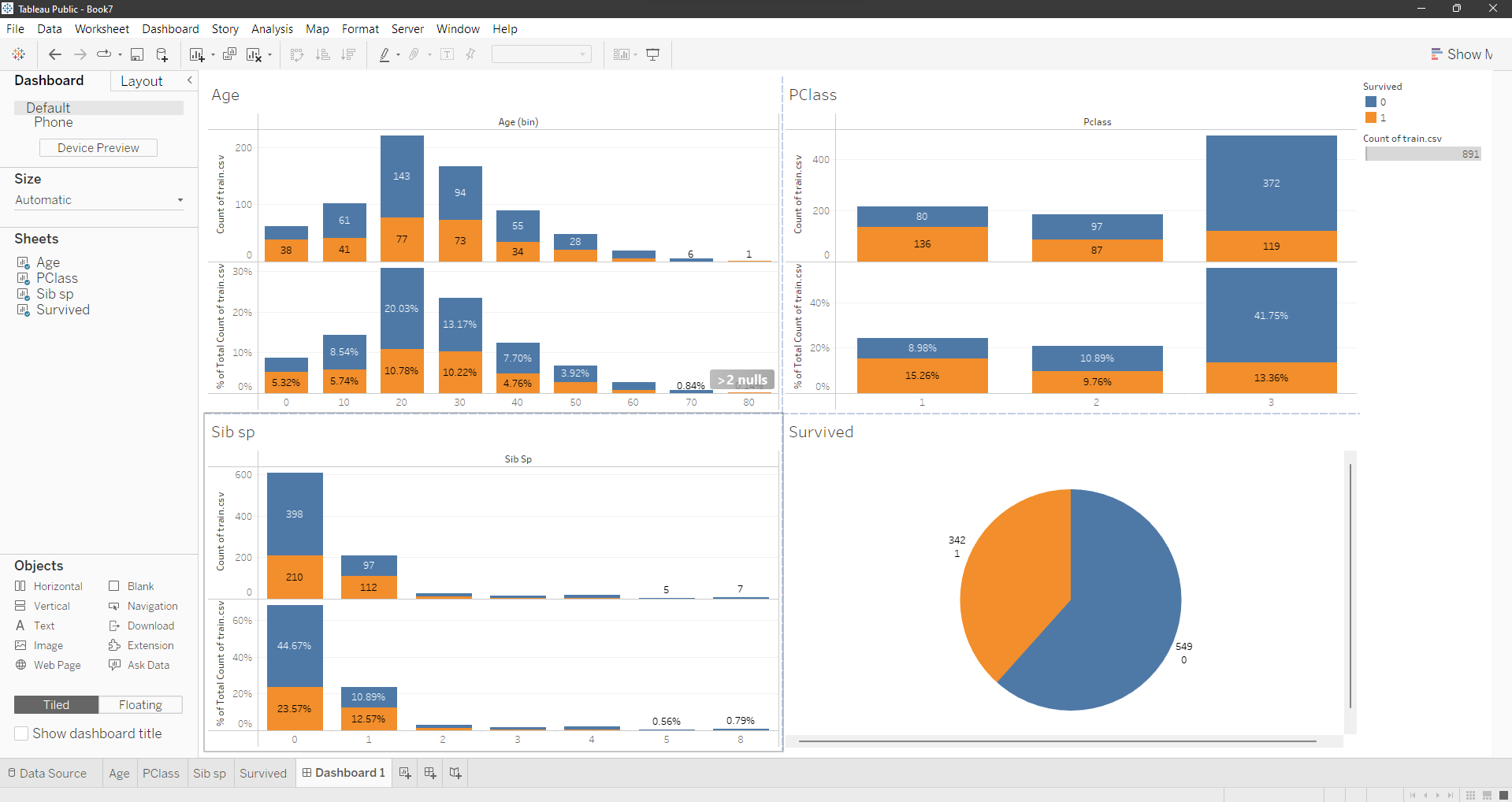
1. Passenger survived out of all passenger



1. Now that our visualization is ready, we move on to Creating Dashboard
2. In the bottom on the right side of Survived cell, there is option for create new dashboard, click on it, you will get this screen



1. Double click on all the sheets that we made and our dashboard is ready



1. To generate a report, save the file, with your tableau public account.
2. This will open up your dashboard on tableau website, Click on download button there, select PDF
3. In include section select Specific sheet from this dashboard and select all the sheets then click on download
4. Generated report is attached below