Project focusing on Data Analysis and Visualization via Tableau

Data Bootcamp – Homework-TABLEAU

kimberly GORDON

2019

Contents

Objectives

Terminology

Requirements

Details

Resources

Objectives

* Aggregate the data in the Citi Bike Trip History Logs
* Build a data dashboard, story , or report
* Work with a timespan of your choosing
* Optional: Merge multiple datasets from different periods

Requirements

* Use Tableau to rapidly manipulate tables of data and create visualizations using a drag-and-drop style interface.
* Connect various data formats such as CSV and Excel Workbooks to Tableau.
* Perform exploratory data analysis using Tableau.
* Create groups and sets.
* Create maps and use built-in U.S. Census data (if applicable).
* Create custom calculations.
* Understand LOD calculations.

Questions answered during the analysis

1. How many trips have been recorded total during the chosen period?

|  |  |
| --- | --- |
| Month | Total Trips |
| January | 300,400 |
| February | 224,736 |
| March | 439,117 |
| April | 670, 780 |
| May | 866,117 |
| June | 936,880 |
| July | 968,842 |
| August | 963,489 |
| September | 953,887 |
| October | 828,711 |
| November | 529,188 |
| December | 399069 |
| Total Trips | 8081216 |
|  |  |

1. By what percentage has total ridership grown?

The percentage of ridership increased the most in March over 90%.

1. How has the proportion of short-term customers and annual subscribers changed?

There are more customers than subscribers.

1. What are the peak hours in which bikes are used during summer months?

The peak hours are between 5pm to 7pm.

1. What are the peak hours in which bikes are used during winter months?

The peak hours are between 5pm to 7pm.

1. Top 10 stations in the city for starting a journey?

8th Avenue and W 31st street – 100,498

Lafayette Street & E 8th Street – 86,692

E 17st & Broadway – 80,166

1. What are the top 10 stations in the city for ending a journey? (Based on data, why?)

E 17st & Broadway – 85,494

8 Ave & W 31 St – 81,367

Lafayette Street & E 8th Street – 84,029

1. Today, what is the gender breakdown of active participants (Male v. Female)?

There is significantly more men riding than women…more than two times the amount of male riders to female riders.

1. How effective has gender outreach been in increasing female ridership over the timespan?

The gender outreach has been very helpful to the point that more women join cycling monthly.

1. How does the average trip duration change by age?

16 year old and 84 year old riders show to ride more than all others More research is needed to research this phenomena.

1. What is the average distance in miles that a bike is ridden?

On the average in 2014, most riders covered a little more that a mile for each trip.

1. Which bikes (by ID) are most likely due for repair or inspection in the timespan?

The bikes that are most like due for repair are the miles that are available in the top 10 popular stations where the journey begins.

1. How variable is the utilization by bike ID?

Most bikes are used for trips that average a little over a mile.

1. City officials would like to see the following visualizations:
   * A static map that plots all bike stations with a visual indication of the most popular locations to start and end a journey with zip code data overlaid on top.
   * Find at least two unexpected phenomena in the data and provide a visualization and analysis to document their presence.
     1. 16 year old and 84 year old riders show to ride more than all others More research is needed to research this phenomena.
     2. Citi ridership is an area where more women need to be involved.

Terminology

**Alias -** An alternative name that you can assign to a field or to a dimension member.

**Bin -** A user-defined grouping of measures in the data source.

**Bookmark -** A **.tbm** file in the Bookmarks folder in the Tableau repository that contains a single worksheet. Much like web browser bookmarks, **.tbm** files are a convenient way to quickly display different analyses.

**Calculated Field -** A new field that you create by using a formula to modify the existing fields in your data source.

**Crosstab -** A text table view. Use text tables to display the numbers associated with dimension members.

**Dashboard -** A combination of several views arranged on a single page. Use dashboards to compare and monitor a variety of data simultaneously.

**Data Pane -** A pane on the left side of the workbook that displays the fields of the data sources to which Tableau is connected. The fields are divided into dimensions and measures. The data pane also displays custom fields such as calculations, binned fields, and groups. You build views of your data by dragging fields from the data pane onto the various shelves that are a part of every worksheet.

**Data Source Page -** A page where you can set up your data source. The data source page generally consists of four main areas − left pane, join area, preview area, and metadata area.

**Dimension -** A field of categorical data. Dimensions typically hold discrete data such as hierarchies and members that cannot be aggregated. Examples of dimensions include dates, customer names, and customer segments.

**Extract -** A saved subset of a data source that you can use to improve performance and analyze offline. You can create an extract by defining filters and limits that include the data you want in the extract.

**Filters Shelf -** A shelf on the left of the workbook that you can use to exclude data from a view by filtering it using measures and dimensions.

**Format Pane -** A pane that contains formatting settings that control the entire worksheet, as well as individual fields in the view. When open, the Format pane appears on the left side of the workbook.

**Level Of Detail (LOD) Expression -** A syntax that supports aggregation at dimensionalities other than the view level. With the level of detail expressions, you can attach one or more dimensions to any aggregate expression.

**Marks -** A part of the view that visually represents one or more rows in a data source. A mark can be, for example, a bar, line, or square. You can control the type, color, and size of marks.

**Marks Card -** A card to the left of the view, where you can drag fields to control mark properties such as type, color, size, shape, label, tooltip, and detail.

**Measure -** A field that is a dependent on value of one or more dimensions. Tableau treats any field containing numeric (quantitative) information as a measure.

**Pages Shelf -** A shelf to the left of the view that you can use to split a view into a sequence of pages based on the members and values in a discrete or continuous field. Adding a field to the Pages shelf is like adding a field to the Rows shelf, except that a new page is created for each new row.

**Rows Shelf -** A shelf at the top of the workbook that you can use to create the rows of a data table. The shelf accepts any number of dimensions and measures. When you place a dimension on the Rows shelf, Tableau creates headers for the members of that dimension. When you place a measure on the Rows shelf, Tableau creates quantitative axes for that measure.

**Shelves -** Named areas to the left and top of the view. You build views by placing fields onto the shelves. Some shelves are available only when you select certain mark types. For example, the Shape shelf is available only when you select the Shape mark type.

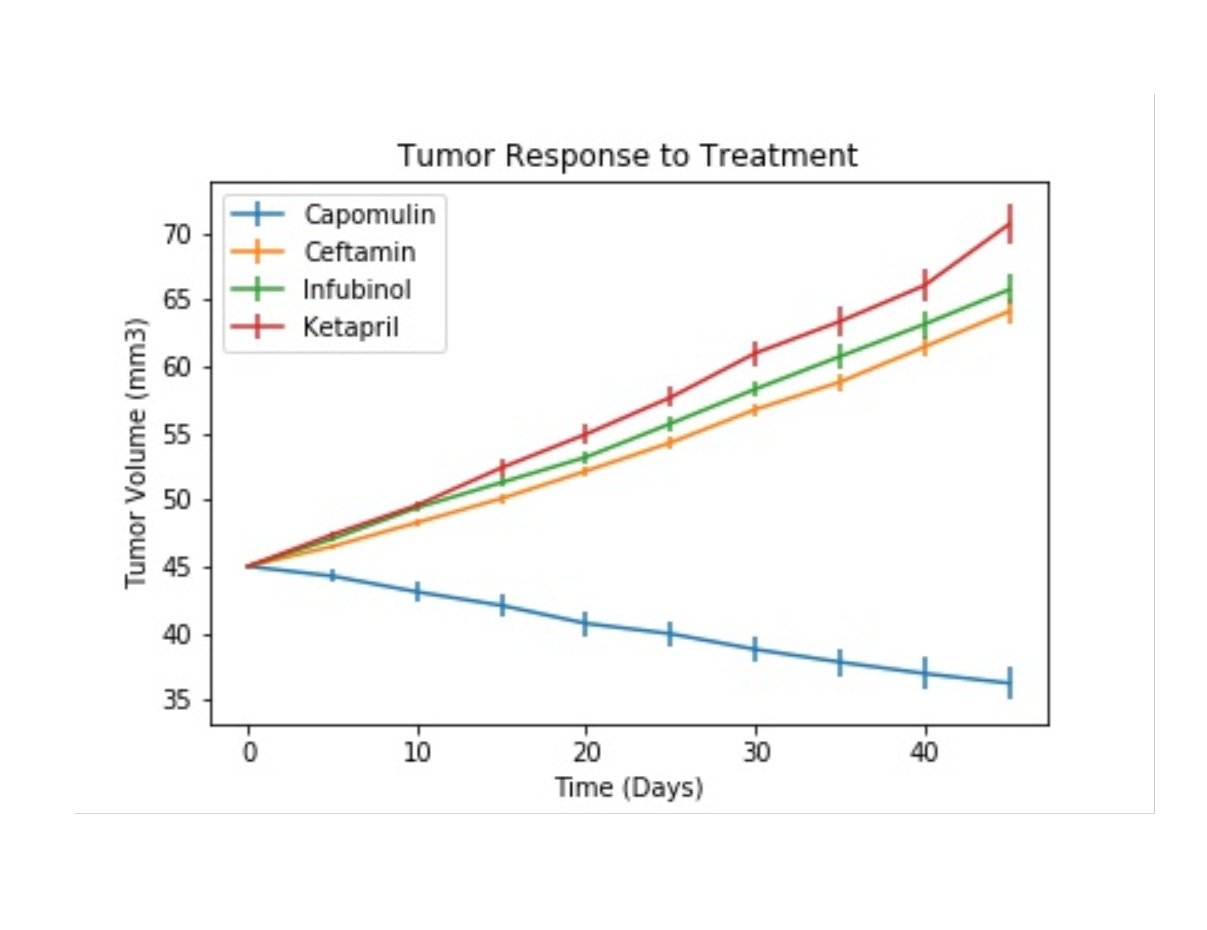
**Tableau -** A business intelligence software that allows anyone to connect to respective data, and then visualize and create interactive, sharable dashboards.

**Workbook -** A file with a **.twb** extension that contains one or more worksheets (and possibly also dashboards and stories).

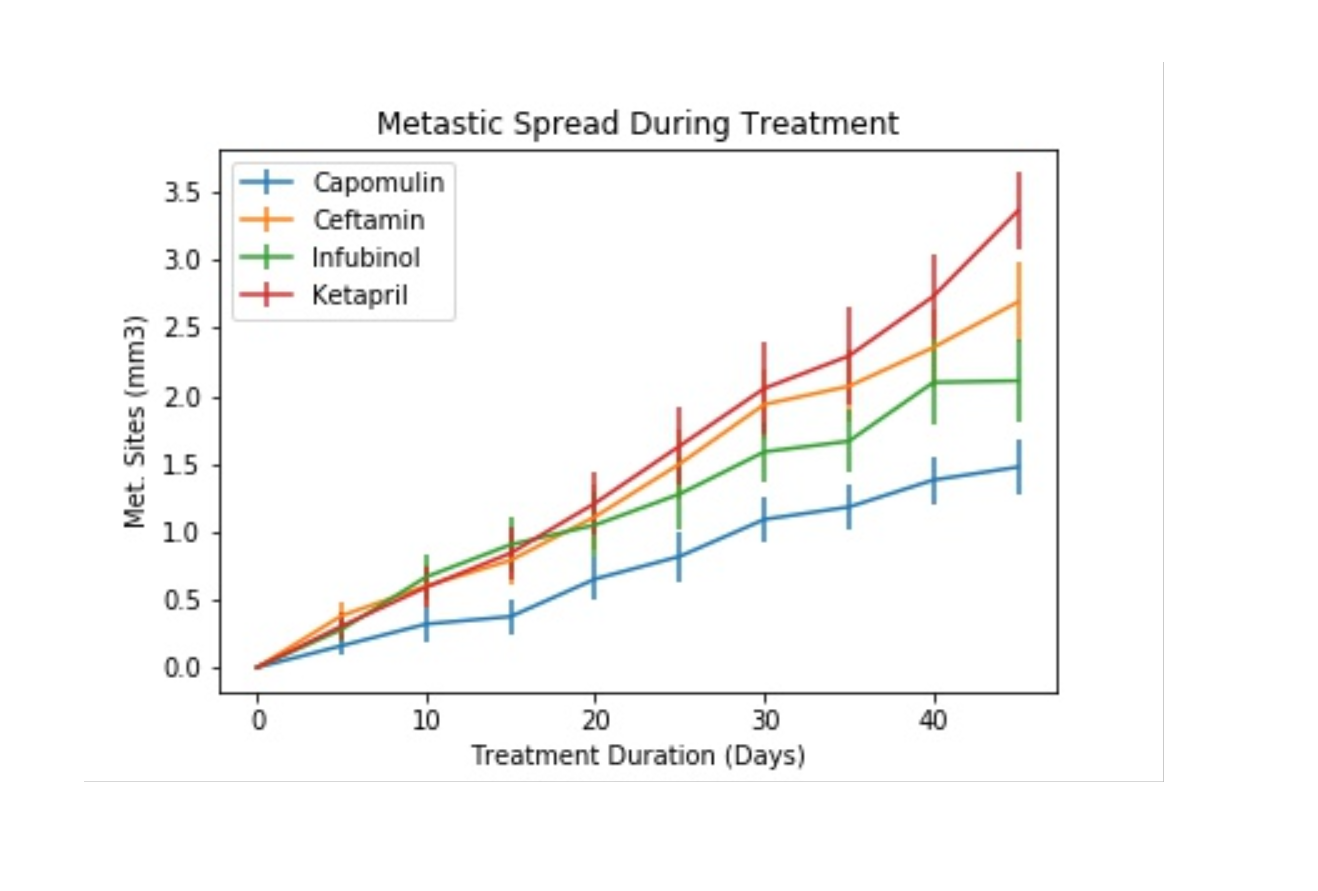
**Worksheet -** A sheet where you build views of your data by dragging fields onto shelves.

Details

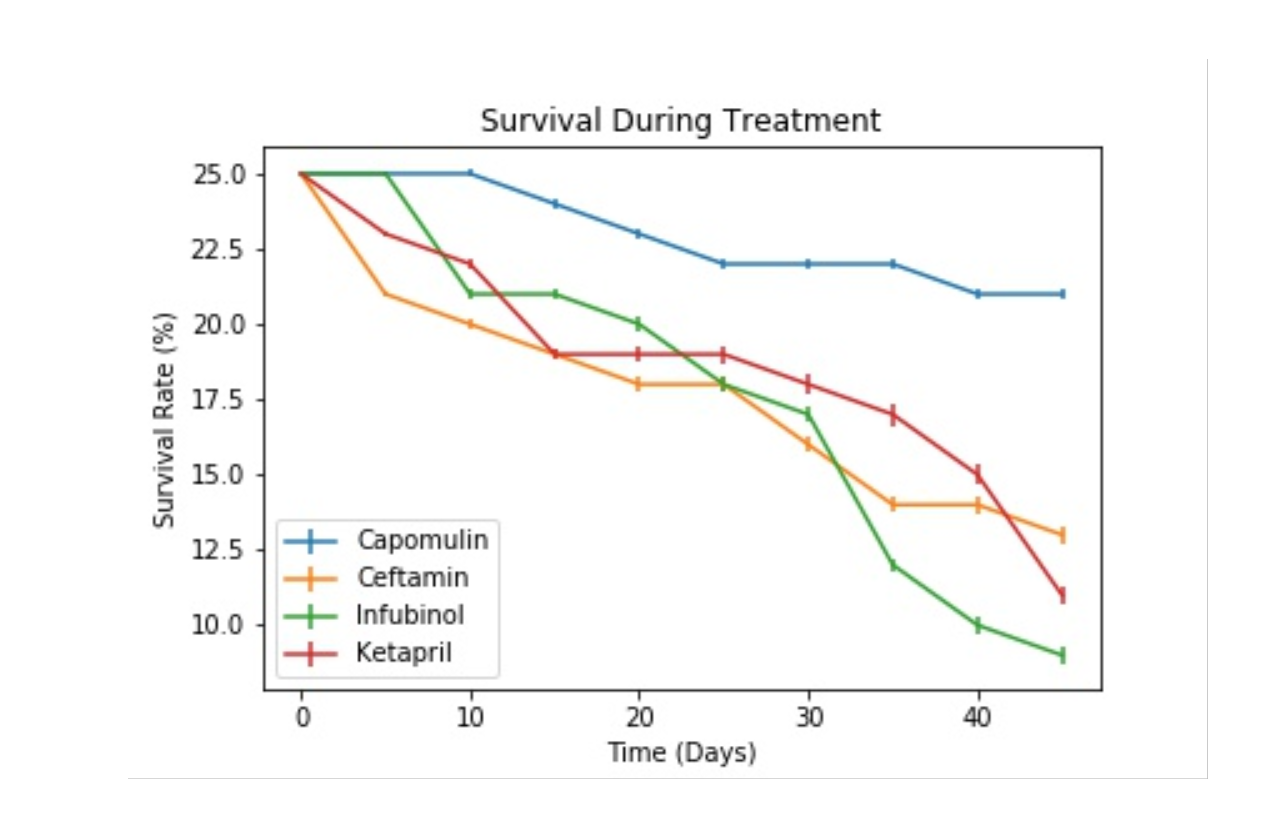
1. Scatter plot that shows how the tumor volume changes over time for each treatment.



1. Create a scatter plot that shows how the number of metastatic (https://en.wikipedia.org/wiki/Metastasis) (cancer spreading) sites changes over time for each treatment.



1. Create a scatter plot that shows the number of mice still alive through the course of treatment (Survival Rate)



1. Create a bar graph that compares the total % tumor volume change for each drug across the full 45 days.

