Case Study

Dataset to use

* VW\_SALES\_DIST\_DETAILS Excel file
* product.txt of FDA Orange Book (OB). please find background information of Orange Book and file description here, [https://www.fda.gov/drugs/drug-approvals-and-databases/orange-book-data-files](about:blank).

Problem

USP produces Reference Standards (RS) in medicine manufacturing to ensure the availability of quality and safe medicine. We produce RS in different categories for global market. Our global leadership needs to understand how RS sales in India have trended since fiscal year 2016, by item category. They also want to see item category revenue as a percent of total, by fiscal year. Furthermore, they want to see sales for the top 10 items and the percent of total sales they represent. Last, they are interested in knowing whether there is a relationship between RS sales (Small Molecules only) and drug applications in the US.

1. Please conduct exploratory data analysis on the VW\_SALES\_DIST\_DETAILS
2. Put together visualizations that will answer the above
3. Discuss insights that you’ve uncovered

Note:

* You are required to conduct data analysis using Python.
* Sales are measured in both units and revenue
* To explore the relationship between RS sales and drug application, you need establish a relationship between “item name” and “Ingredient” in OB product. Please use your creativity and reasonable assumption to map “item name” with “ingredient”.