MLDS 411
DATA VISUALIZATION

Winter 2024 Labs

Alice Zhao

Northwestern | McCORMICK SCHOOL OF ENGINEERING

Schedule

Week	Date	Time	Tableau Topics	Python Topics
1	January 12	11am	Tableau Review	Library Overview
2	January 19	12pm	Joining, Blending	Plotly, Dash
3	January 26	11am	Groups, Hierarchies, Sets	Bokeh
4	February 2	11am	Parameters, Tool Tips	Wordcloud
5	February 9	11am	Dashboards, Stories & Actions	NetworkX
6	February 16	11am	Maps	GeoPandas, Folium
7	February 23	11am	Calculated Fields, Calculations	Plotnine
8	March 1	11am	Forecasting, Clustering	Python + Tableau

Lab 1 Questions

- Range formatting
  - Measure >> Default Properties >> Number Format
- Map formatting
  - Map >> Background Layers
- Color palette adjustability
  - Default is the palette is locked
  - Can potentially adjust it using a calculated field and a dual axis

Today's Theme: Combining Data

Unions and Joins
- Identical to the concepts in SQL and Python
- Done as a first step on the Data Source tab

This is my go to as a data scientist who learned Tableau before the early 2020's

Blending
- Tableau-specific feature introduced in 2010
- Allows you to combine data sources on each sheet instead of the entire workbook

These are less commonly used, but you may encounter them in other workbooks

Relationships
- Tableau-specific feature introduced in 2020
- A more flexible and smarter version of joins

This is the Tableau-recommended way of combining data and the default

Demo: H&M Sales

1. Data Cleaning

- Data: H&M Sales
- Demo:
  - Split
  - Hide
  - Calculated Field
  - Group
  - Filter

2. Union

- Data: H&M Sales
- Demo:
  - Union the tabs together
  - Sales by Category by Year
  - Overlay the Sum of Sales text

3. Blend

- Data: H&M Sales 2018 & 2019
- Demo:
  - Blend the sheets together
  - Sales by Category by Year
  - **Blue** vs **Orange** colors
  - Overlay the Sum of Sales text

Demo: Book Shop Database

1. Join

- Data: Books, Info
- Demo:
  - Join books and info
  - View data tab
  - Flip through join types
  - View the book count

2. Relationship

- Data: Books, Info
- Demo:
  - Create a books and info relationship
  - Noodle them together
  - View the book count

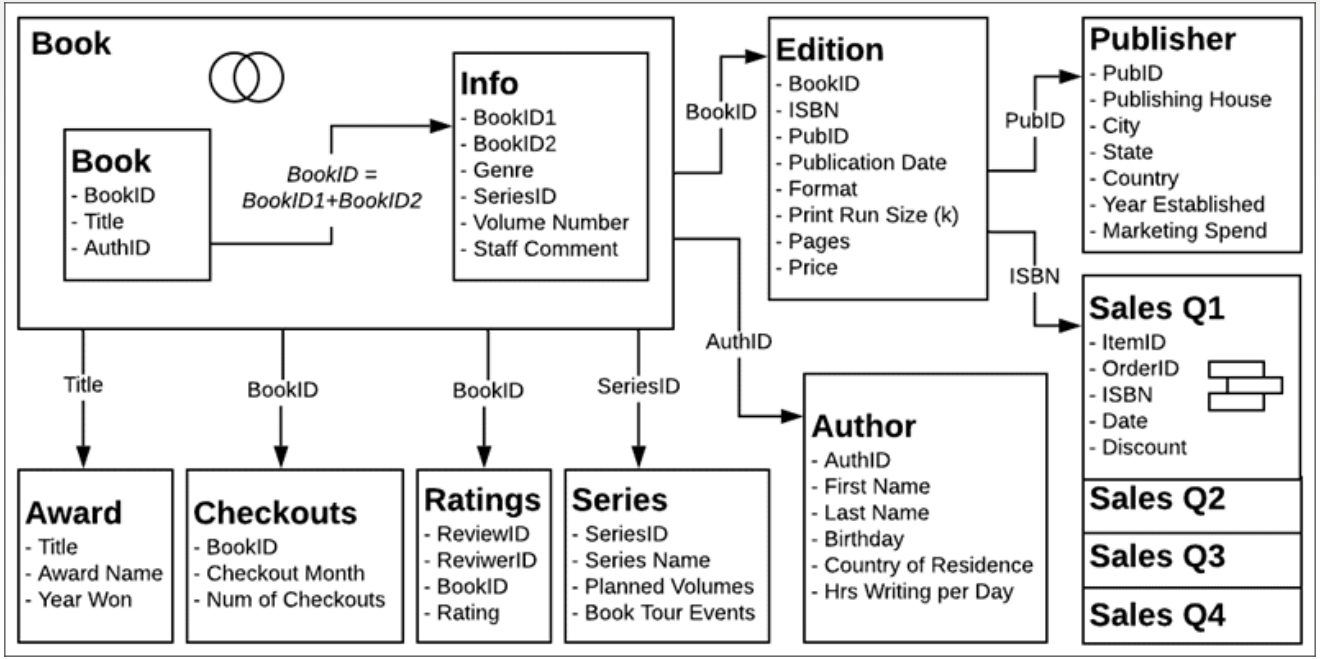
3. Join

- Data: Author, Rating
- Demo:
  - Average Author Rating by Last Name
  - View the book count

4. Relationship

- Data: Author, Rating
- Demo:
  - Average Author Rating by Last Name
  - View the book count

Book Shop Data Model



https://help.tableau.com/current/pro/desktop/en-us/bookshop_data.htm

Exercise: Book Shop Database

- Who is the author with most books published?
  - Do this using relationships
  - Do this using joins

- In which month of the year are the fewest books published?
  - Do this using relationships
  - Do this using joins

Combining Data Summary

1. Start with Relationships

- Default option in Tableau
- Good for complex data models, tables with different row granularities and non-technical users

2. Join if you're more comfortable

- Good for simple data models, if you want more control over your join types and you're more familiar with SQL

3. Don't use Blending unless you have to

- Data is specific to each sheet, only allows for left joins, etc.

Data Visualization in Python Overview

Standard Libraries

pandas

matplotlib

seaborn

Mapping

GeoPandas

Folium

Interactive Visualizations

plotly | Dash

bokeh

Specialty

amueller / word_cloud

NetworkX
Network Analysis in Python

plotnine

Plotly and Dash

Plotly

- Plotting library that allows you to create interactive plots
- Example plotly charts: https://plotly.com/python/basic-charts/
- There is also an “express” version of plotly, which has simplified syntax

Dash

- Allows you to create web applications with interactive plotly plots

Demo

- Candy_Analysis_and_Visualizations.ipynb
- app.py