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 & Relationships	Plotly, Dash
3	January 26	11am	Hierarchies, Groups & Sets	Bokeh
4	February 2	11am	Filters, Parameters & Tooltips	Wordcloud
5	February 9	11am	Dashboards, Stories & Actions	NetworkX
6	February 16	11am	Maps	GeoPandas, Folium
7	February 23	11am	Calculated Fields, Table Calculations	Plotnine
8	March 1	11am	Forecasting, Clustering	Python + Tableau

Data Visualization Final Project

- Feedback provided via Canvas
- Final presentations on Monday, March 11
  - Rebeca will be grading the presentations
  - I will be grading the Tableau dashboards
- Grading (out of 10)
  - 5: Basic dashboard
  - 6-9: Nice aesthetic, includes interactivity, multiple data sources, etc.
  - 10: Tableau Public gallery quality

Lab 7 Follow Ups

Will cover near the end of lecture:

- Can you change the color using a calculated field?
- Data is not updating with refresh
- How to drill down in Tableau
- How to pivot / melt within the Data Source tab
- How to remove the Abc column

Today's Theme: Modeling

Clustering

- K-Means Clustering

Statistics

- Constant Lines
- Trend Lines

Forecasting

- Additive vs Multiplicative Models
- Comparing Actuals vs Predictions

Clustering

- **K-Means Clustering**
  1. Select two fields (CO2 vs GDP)
  2. Scale the data (optional in Tableau)
  3. Fit a K-Means model (no random state, includes scaling)
  4. Modify the number of clusters

- **In Practice**
  1. Play around with the features and k values in Tableau
  2. Do the actual coding within Python

Statistics

- **Constant Lines**
  - Profit by Month: Average, Zero and Minimum
- **Trend Lines**
  - Sales by Quarter
    - Trend Line Types: Linear, Exponential, Polynomial, etc.
    - Color by Category

Forecasting

- **Exponential Smoothing**
  - Forecast Sales by Quarter, Sales by Month
  - Forecast >> Forecast Options
    - Additive vs Multiplicative Models
    - Forecasting_Demo.ipynb

- **Comparing Actuals vs Predictions**
  1. Forecast Sales by Month
  2. Copy predicted data points into new worksheet >> Cmd+A >> Cmd+C >> Cmd+V
  3. Update Order Date and Predicted Sales fields
  4. Create a visualization with both the sales and predicted sales
  5. Choose Dual Axis >> Synchronize Axis

Modeling Exercises

- Clustering
  - Looking at Profit vs Sales by Sub-Category, apply K-Means Clustering
    - What is a good number of clusters?
    - How would you interpret the clusters?
- Trend Lines
  - Looking at the monthly Sales by Country, add Trend Lines
    - What type of trend line would you choose?
    - How would you interpret the trends?
- Forecasting
  - Looking at monthly Sales, predict 2024 Sales
  - Looking at monthly Sales, predict Q3 & Q4 2023 Sales
    - Compare the predicted vs actual sales

Modeling Summary

Analytics Tab

- Summarize
- Model
  - Trend Line
  - Forecast
  - Cluster

In Practice

- Test models out in Tableau >> Implement in Python
- Implement models in Python >> Visualize in Tableau

Data Visualization in Python Overview

Standard Libraries

Mapping

Interactive Visualizations

Specialty

Python + Tableau

TabPy

- Stands for Tableau Python Server
- Developed by Tableau (https://github.com/tableau/TabPy)
- Allows you to write Python code within Tableau

Conclusion

- While you can write Python code within Tableau, my suggestion is to do the analysis within Python, and then import the data into Tableau

Python + Tableau

TabPy Installation

- python -m pip install --upgrade pip
- pip install tabpy
- pip install --upgrade pyarrow

Launch TabPy (in Terminal)

- tabpy

Enable Tableau

- Help > Settings and Performance > Manage External Service Connection > TabPy
- Hostname: http://localhost:9004/
- Port: 9004

Python + Tableau

Simple Example

- Returns a Boolean value of which profits are positive

```
SCRIPT_BOOL("
lst= []
for i in _arg1 :
       lst.append(i>0)
return lst",
SUM([Profit]))
```

Complex Example

- DBSCAN and real time predictions
- https://www.tableau.com/blog/building-advanced-analytics-applications-tabpy

Lab 7 Follow Ups

Will cover near the end of lecture:

- Can you change the color using a calculated field?
- Data is not updating with refresh
- How to drill down in Tableau
- How to pivot / melt within the Data Source tab
- How to remove the Abc column

Lab 7 Follow Ups

- Can you change the color using a calculated field?
  - Short answer: not really, but there is a lengthy workaround
  - https://community.tableau.com/s/question/0D54T00000C6hH2SAJ/hex-values

- Data is not updating with refresh
  - Try refreshing both the data and the dashboard
  - Try clearing your cache (Help >> Settings and Performance >> Clear Cache)
  - https://www.reddit.com/r/tableau/comments/13ldc1i/tableau_dashboard_not_refreshing_with_an_extract/

Lab 7 Follow Ups

- How to drill down in Tableau
  - This may not be possible
  - Most promising alternatives: actions, creating checkboxes of levels
- How to pivot / melt within the data source tab
  - Select multiple columns >> Right click >> Pivot
- How to remove the Abc
  - Option 1: Marks >> Polygon
  - Option 2: Drag second copy of last pill in Rows over Text >> Uncheck Show Header

Why Use Tableau?

- **Connect to a variety of data sources**
  - Joining, Blending & Relationships
- **Create interactive visualizations**
  - Filters, Parameters & Tooltips
  - Dashboards, Stories & Actions
  - Maps
- **Perform basic to advanced analysis**
  - Hierarchies, Groups & Sets
  - Calculated Fields, Table Calculations
  - Forecasting, Clustering

Data Visualization Options

Difficulty	Category	Software	Details
1	Fundamental Tool	Excel	Most popular and widely used
2	Business Intelligence Tools	Tableau PowerBI Looker Qlik	Often used for interactive visualizations at large companies
3	Programming Languages	Python R	Many available data visualization libraries
4	Web Development Frameworks	D3.js	Allows for creation of custom visualizations