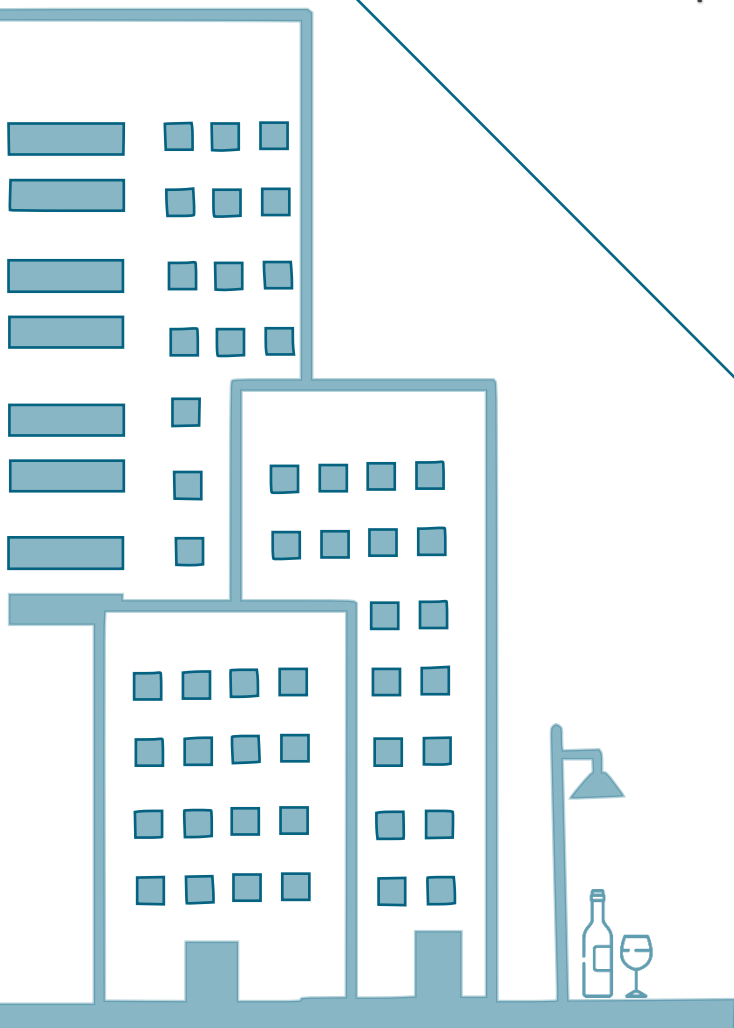# Data Management and Visualization Assignment Report

Iowa State Liquor sales and Volume Dashboard

November 20, 2021

Prepared By

Shubham Rai

Sarthak Taru

Group: *Pair M*

Abstract

The rate of binge drinking in Iowa is among the highest in the United States.

Data Collection

We have collected the Iowa state liquor data from Iowa state government site (https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy). This data is provided by Iowa Department of Commerce, Alcoholic Beverages Division and can be downloaded in CSV format. This dataset contains every individual purchase of liquor in the state of Iowa from retailers since 2014. Each row from dataset contains the individual purchase of the liquor bottle with store number, store location, liquor type, volume, brand etc. The dataset contains almost **8.14 million records** before cleaning and pre-processing. Space utilisation of CSV file is 3.17 GB. This covers the **volume aspect of big data**. We are using this data set to analyse total spirits sales in Iowa of over the years from 2018 to October 2021.

Data Exploration, pre-processing, Cleaning and Transformation

Pre-processing and Exploration

Initial Data Exploration and pre-processing was **performed in Python** with goal to explore fields, reformatting geographic fields and removing nulls. Initially, we explored data types of the field and checked for null values for each field and dropped rows which had null values for fields that we were going to analyse. Later we reformatted Geographical fields to desired format. This file was exported to be cleaned and transformed in Tableau Prep.
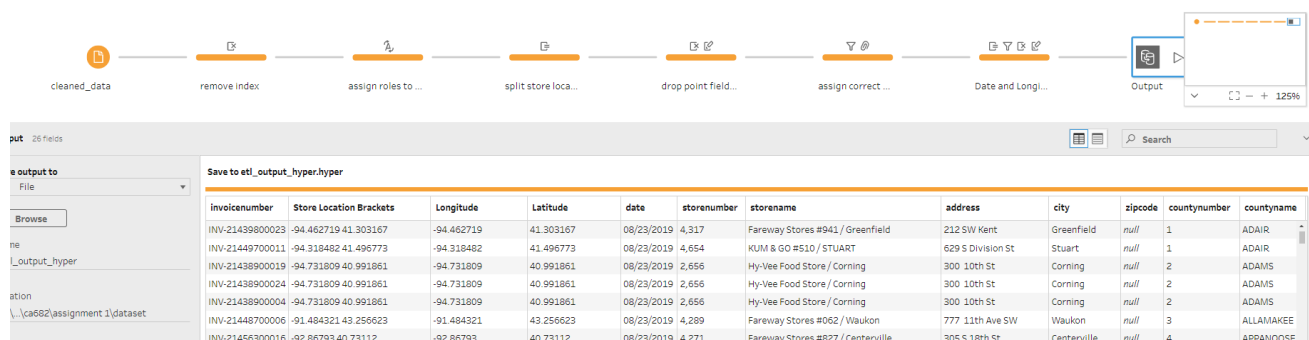
Cleaning and Data transformation

For **cleaning and Data transformation** we used **Tableau Prep** due to its easy-to-use approach and ability to export data in hyper file.

Steps Performed

1) Drop unwanted fields
2) Creating Regions field based on city field to divide Iowa state in geographic regions.
3) Creating calculated field to assign geographic roles to be used in Tableau Reporting.
4) Create calculated field to get State profit based on retail and sales field.
5) Data extraction for date range between 2018 and 2021
6) Export dataset in **Hyper File instead of CSV** as hyper file is in-memory data engine and **optimised for analytical query processing**.

Outcome of above steps was Hyper file to be used in Tableau for Reporting.

Screenshot Tableau Prep

Visualisation

For Dashboard Design reference, we followed *Visual Analysis Best Practices: A Guidebook* **whitepaper** from Tableau.

Purpose

We aimed to create a dashboard for IOWA state employees to get all necessary information about Sales, Profit and Volume for Liquor sold in single view.

Iowa State
Liquor Volume and Sales Dashboard

| Total Bottles Sold	State Profit	Top Selling Category	Top Vendor
11,007,913	$3,849,254	American Vodkas	DIAGEO AMERICAS
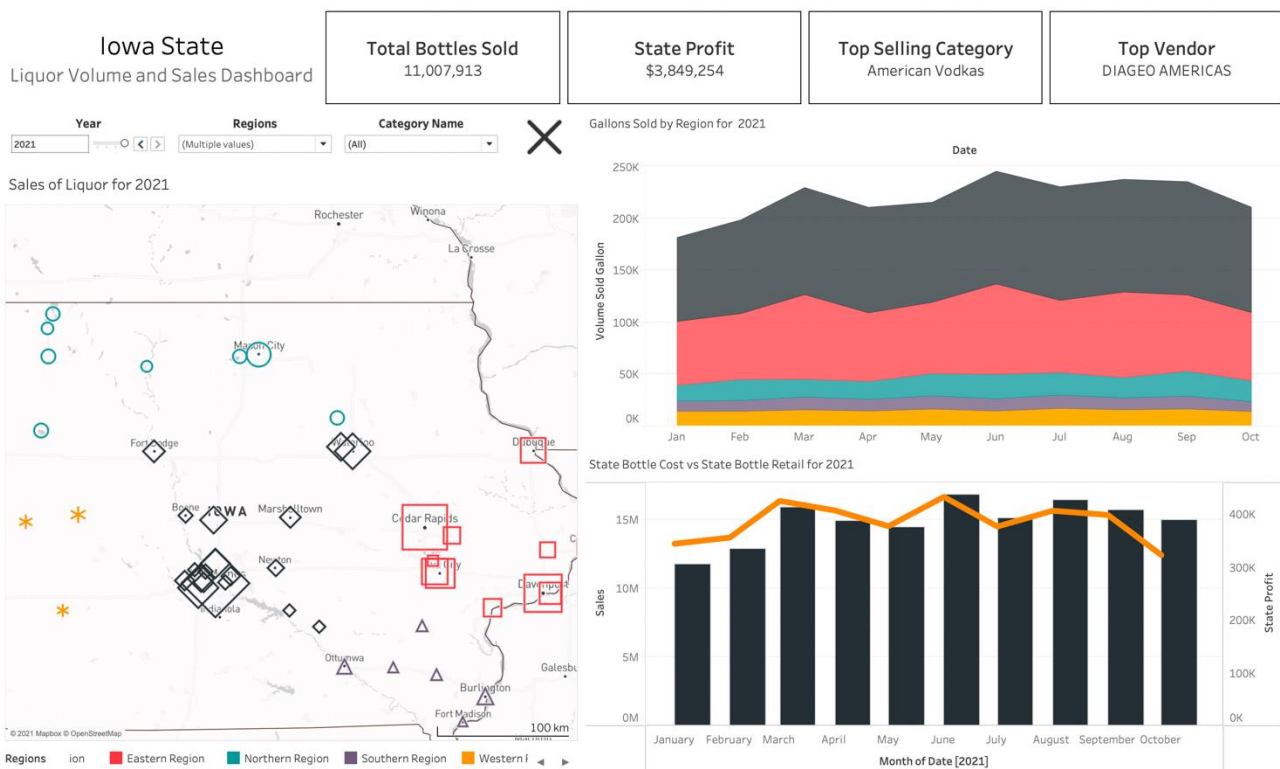
Design Choices

Creating effective Views

1) **Avoided overloading views**: We have tried to keep the dashboard as minimalistic as possible. Only relevant mark types were included to be displayed. Also, not all users will need filters to be displayed while printing or exporting dashboard as report, hence we have added close button to hide filters.
2) **Limit the number of colours and shapes in a single view**: We tried to limit the number of marks on the charts. Limited number of colours were chosen to keep the dashboard clean and minimalistic.
3) **Orient views for legibility**: Most of the dashboard labels are kept horizontally oriented to provide ease to users while reading.

Colour Choices

1) **Primary Theme**: The main theme we chose for our dashboard is black and white with custom colour palate for regions. Symbol map displayed was created with custom colour theme with help of MapBox to match overall theme
2) **Colour Palette:** We used **custom colour palette** for our dashboard to create contrasting colours. We have considered colour blind audience while choosing

colours and used online tools to visualise our dashboard in **simulated colour-blind environment**. All the colours were distinguishable in Area Chart. Screenshot to **simulate dashboard for colour blindness**
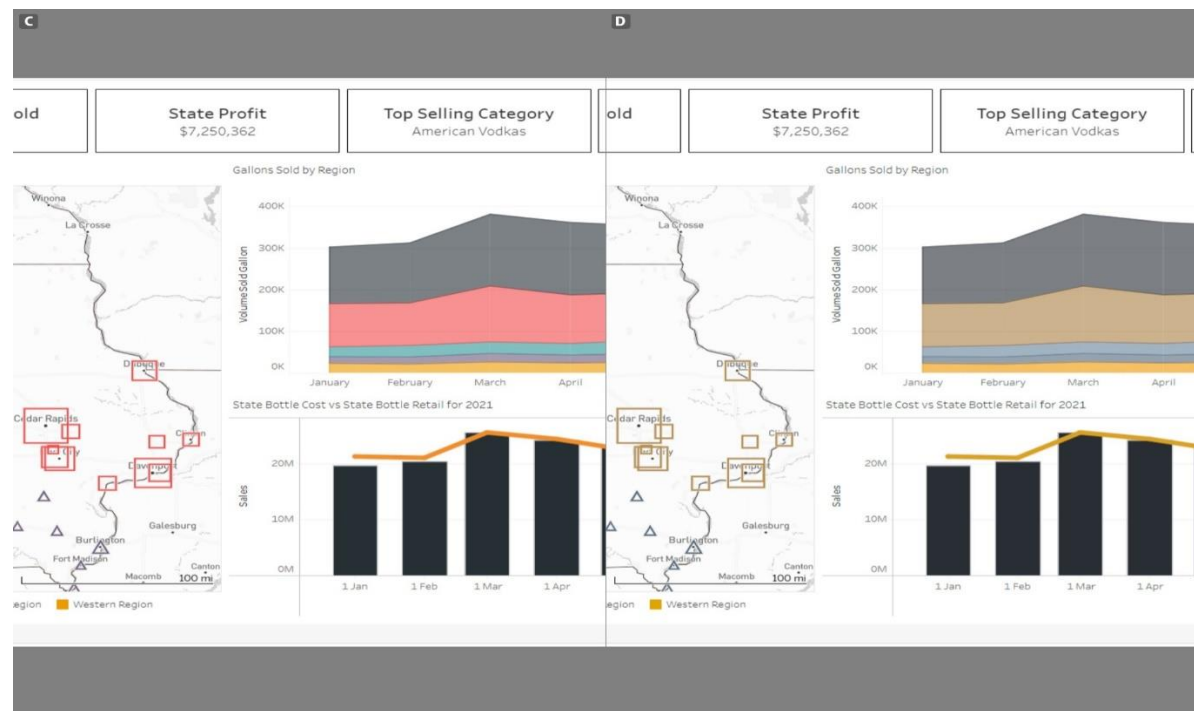
Animation Choice: We have added animations which is in sync with year slider filter to show smooth transitions and users can notice changes while applying filter. Also animations were added while filtering data when choosing city mark on map chart.

Choosing Chart type

1) **Sales Symbol Map:** We created a Symbol Map to highlight sales based on region. **Marks** on chart **indicates cities** and size of marks grows and decrease based on sales made in city. **Custom designed map** was used which was created with help of MapBox. We also used **shape card** in Tableau to distinguish between different regions as it will be easier for **colour blind** audience. Colour choices were also made considering colour blind audience. To view sales by year we have provided **Year Filter** for user interaction along with **search button to filter by city**. Besides above we have also added **zoom in/out** ability to users **for cases of overlapping marks**.

2) **Area Chart:** We chose **area chart** to show progression of **volume of liquor sold** in gallons **over time**. We have used colour card to distinguish different regions in chart.

3) **Dual Axis chart:** To provide comparison of sales vs state profit over time, Dual axis was the best choice. Dual axis charts provide an easy-to-read chart and compares 2 measures effectively over time. X axis represents time progression for each year.

4) Additionally, we have also added **boxes that showcases important aspects** such as **top seller, bottles sold, total sales and top selling category**.

Tools and Libraries used

Data Pre-processing and Exploration: We have used Python to perform initial pre-processing and Data exploration.

Data Cleaning and Transformation: Tableau prep was used for cleaning, transformation and filtering out data.

Visualisation: Tableau 2021.3 was used for creating dashboard and charts.

Colour Blind simulation: Pilestone.com was used to simulate dashboard for colour blindness.

Custom Map: MapBox was used to create custom Symbol Map to be integrated with Dashboard.

The whole dashboard is designed in Tableau Desktop and published to Tableau server https://public.tableau.com/app/profile/shubham.rai7724/viz/IowaSalesConsuptionDashboard/Dashboard?publish=yes

Conclusion

Dashboard was created with aim to provide IOWA state employees to glance through sales and volume sold figures by year. Most of the time was spent in colour choices and dashboard creation. The charts are simple, easy to read, and interactivity is added in each chart with addition filters. The most interesting observation in sales chart was although each year October had highest sales of Liquor, but state profit did not grow in ratio with sales which is case with other months. We can also see gradual increase in sales and volume of alcohol sold from January until December. Most probably the sales of liquor are high due to holiday season. Highest revenue generation is from Central Region of IOWA state and sales of alcohol is highest in capital city of Des Moines. Top Vendor in IOWA state in all regions is Diageo Americas. For future development, A YOY and MOM chart can be added to get better comparison between dates.

References

Tableau Visual Guidebook whitepaper

For integration of custom maps style we followed following tutorial
MapBox style in Tableau