



Cardiff Metropolitan University

Cardiff School of Technologies

Academic Year: 2025-2026

Module: DAS7000 – Data Analytics & Visualization

Module Leader: Dr Sandeep Singh

ABC License Endorsements Data Analysis Using Python and Tableau

Student Name: Jamal Salem Amro

Student ID: 20068275

MSc Programme: MSc Data Science

Date of Submission: 18/12/20

Table of Contents

1. Introduction	4
2. Python Data Analysis Summary	4
2.1 Data Preprocessing and Feature Engineering.....	4
2.1.1 Data Cleaning.....	4
2.1.2 Feature Engineering	4
2.2 Exploratory Data Analysis (EDA)	5
3. Tableau Dashboard and Summary of Insights	6
3.1 Overview of Dashboard	6
3.2 Key Insights	7
3.3 Storytelling	7
5. Conclusion	8
6. References	9

List of Figures

Figure 1. Distribution of License Statuses	5
Figure 2. Distribution of business license categories and sub-categories	6
Figure 3. Monthly trend of expiring licenses	6
Figure 4. ABC License Overview Dashboard	7

1. Introduction

This study analyses Alcoholic Beverage Control (ABC) license endorsements dataset published by the Louisville Metro KY Government (Louisville Metro Government, 2024). The ABC License Endorsements Database includes information about all businesses with Alcoholic Beverage Control licenses issued by Louisville Metro Government. This database includes endorsement types held by the different businesses, their current licenses status, where they are located geographically, and when their licenses will expire, etc. Therefore, this dataset lends itself to exploratory analysis as well as visualization methods. There are also data quality issues, e.g. missing values, having different license statuses, that provide for meaningful data preprocessing and analysis. Through this database it is possible to conduct a very technical analysis and to generate practical, business-type insights that align with the purpose of this study.

2. Python Data Analysis Summary

2.1 Data Preprocessing and Feature Engineering

2.1.1 Data Cleaning

The first stage in the analysis was preparing the data. The dataset contained many errors, including missing values and inconsistent labels. All dates were transformed into the correct datetime format, which also allowed for further processes on the dates. All text was standardized by making it lowercase, and removing extra spaces, which helped to improve the consistency of categories. To help load data, to clean it, to manipulate it, to do numerical processing and to manage missing values, I used Python libraries such as Pandas and NumPy. These libraries are widely regarded as essential libraries for scientific computing and data analysis in Python(Sapre & Vartak, 2020).

2.1.2 Feature Engineering

Feature engineering has been done to add depth to the analysis of the cleaned data by allowing temporal, spatial and risk-based analyses to be applied to the dataset.

To allow for compliance and risk assessment, different indicators of license status were created. These indicators include Days Until Expiry, Is Expired, Is Expiring Soon, and Estimated License Age (Est_License_Age). These license indicators provide insight into what the license status of a business is and how close the license is to expiring.

To provide a standardized method for categorizing licenses, Cat Group and Sub Category were created. These categories group together various types of business licenses into a category that is easier to understand.

2.2 Exploratory Data Analysis (EDA)

Exploratory plots were created in Python by using the libraries Matplotlib, Seaborn, and Plotly, all of which provide great opportunities for making static as well as interactive visual representations of data (Sial et al., 2021). The distribution of license type as seen in figure 1 also indicates that many licenses have either expired, or will be expiring soon.

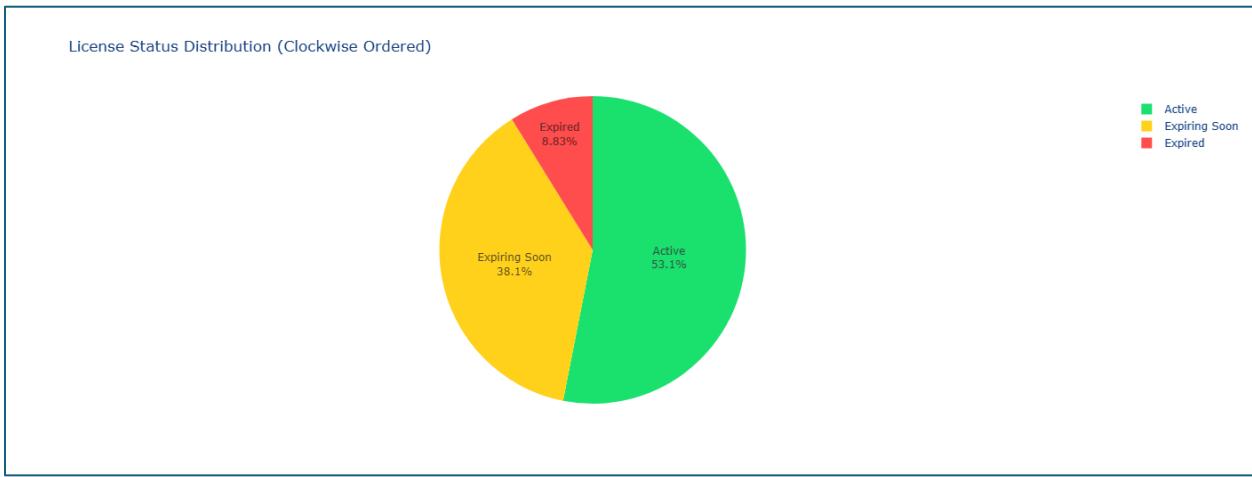


Figure 1. Distribution of License Statuses

The introduction of Category and Sub-Category fields helped organize the dataset into clearer business segments. As shown in figure 2, most licenses fall under a small number of major groups, with Food and Alcohol representing the largest categories.

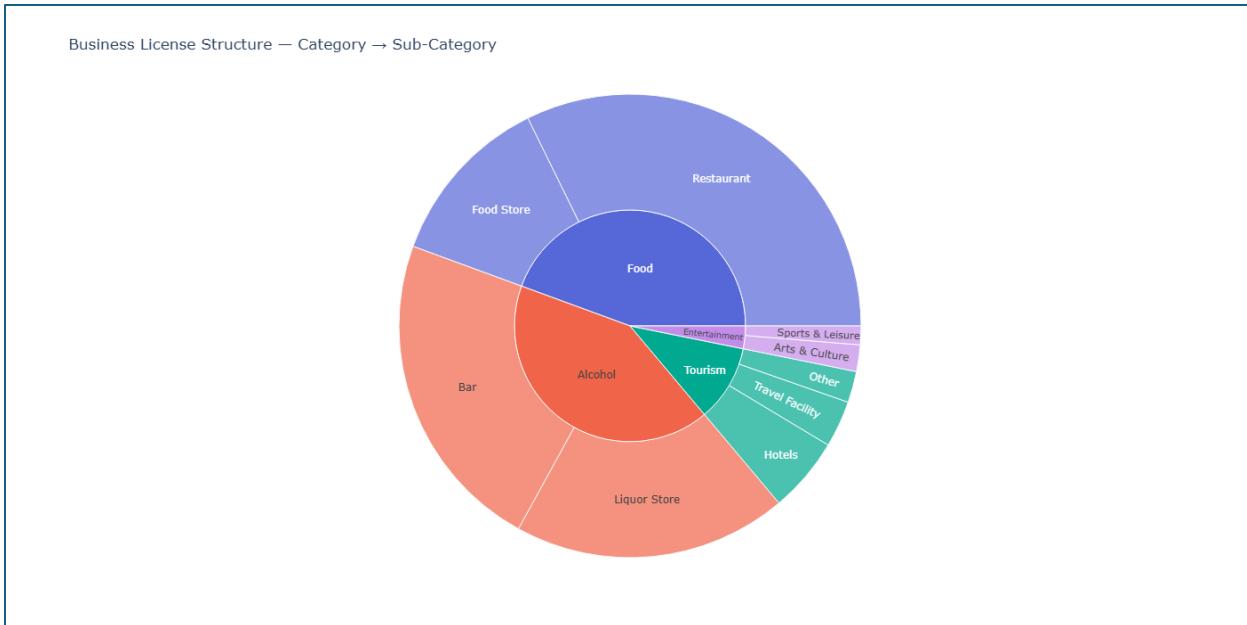


Figure 2. Distribution of business license categories and sub-categories

Certain months have significantly higher license expirations than other months and that administrative cycles will generally coincide with these periods as shown in figure 3.



Figure 3. Monthly trend of expiring licenses

3. Tableau Dashboard and Summary of Insights

3.1 Overview of Dashboard

After data preparation and cleaning, the data was brought into Tableau so users can visually view the data and interactively analyze it. The dashboard features many visual elements such as a map that shows where all businesses are located, key performance indicator (KPIs) to indicate whether businesses are currently licensed, endorsement

charts that show the distribution of endorsements among businesses, and time series charts that show when businesses receive their licenses. In addition, filters have been set up for users to view the data based on license status and category type allowing users flexibility to explore the data and perform their analyses as shown in figure 4.

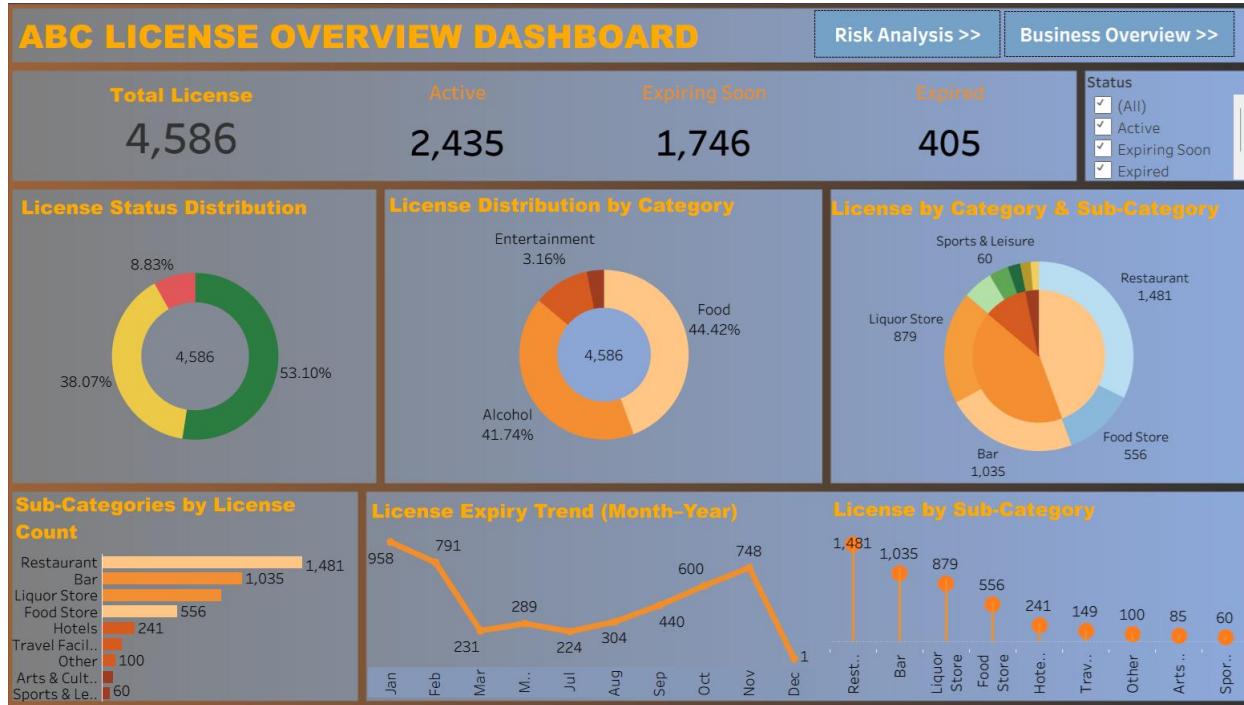


Figure 4. ABC License Overview Dashboard

3.2 Key Insights

Tableau dashboards provide users with the ability to view data in both a spatial and categorical format. Each method allows users to identify meaningful data patterns and trends that facilitate business building and analytical storytelling from data sources to end users (O'brien & Stone, 2020). The distribution of licenses geographically was remarkable, as some areas had significant concentrations of businesses that were licensed. The endorsements were heavily concentrated in a few categories, while there appeared to be moderate seasonal trends with respect to issuing licenses over time. However, based on the analysis of expired licenses and those set to expire soon, there seemed to be substantial numbers of expired and soon to expire licenses, which could represent compliance risks for regulatory bodies.

3.3 Storytelling

The Tableau stories provide the structure for how the insights were laid out logically and guide the user through the different areas: license status, overall businesses activities, and conducting a risk assessment of expired licenses. This approach to storytelling

enhances the user's ability to interpret the insights as they connect each visual component together into an overall story.

5. Conclusion

The report illustrates combining the data analysis capabilities of the Python programming language with Tableau's visualization capabilities provide a method for accurate analysis and effective communication of licensing trends, thereby promoting an evidence-based analysis of distribution and regulatory business activities across Louisville Metro KY.

6. References

- Louisville Metro Government (2024) Louisville Metro, KY ABC License Endorsements. Available at: <https://catalog.data.gov/dataset/louisville-metro-ky-abc-license-endorsements> (Accessed: 07 November 2025).
- O'brien, A. D., & Stone, D. N. (2020). Yes, you can import, analyze, and create dashboards and storyboards in tableau! the gbi case. *Journal of Emerging Technologies in Accounting*, 17(1), 21–31. <https://doi.org/10.2308/jeta-52760>
- Sapre, A., & Vartak, S. (2020). Scientific Computing and Data Analysis using NumPy and Pandas. *International Research Journal of Engineering and Technology*. www.irjet.net
- Sial, A. H. R., Rashdi, S. Y. S., & Khan, A. H. (2021). Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python. *International Journal of Advanced Trends in Computer Science and Engineering*, 10(1), 277–281. <https://doi.org/10.30534/ijatcse/2021/391012021>

I confirm that this assignment is my own work. In accordance with the brief assessment, AI tools were used only for proofreading and language refinement. All data processing, analysis, interpretation, visualizations, and conclusions are my own work and are supported by appropriate academic sources.