**INT375**

**PROJECT REPORT**

(Project Semester January-April 2025)

**Submitted by**

Karman Thethi

Registration No : **12303434**

Programme and Section : **B.Tech (CSE) & K23ED**

Course Code : **INT375**

Under the Guidance of

**Dr. Dhiraj Kapila ( UID: 23509 )**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that **Karman Thethi** bearing Registration no. **12303434** has completed **INT 375** project under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of** CSE

Lovely Professional University

Phagwara, Punjab.

Date: 12-04-25

**DECLARATION**

I, Karman Thethi, student of B.Tech under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 10 / 04/ 2025 Signature

Registration No. **12303434** **Karman Thethi**

**1. Introduction**

With the increasing demand for certified professionals across diverse industries—from cosmetologists to tattoo artists—the process of professional licensing has become a crucial element in workforce regulation and development. This project focuses on analyzing a comprehensive dataset of **professional license counts** across various U.S. counties and professional areas.

Using **Python** and libraries like **Pandas, NumPy, Matplotlib, and Seaborn**, we conducted a detailed **Exploratory Data Analysis (EDA)** to uncover patterns and trends over time. The main objective is to help stakeholders, regulators, and training institutions understand the data through EDA

**2. Source of dataset**

https://data.wa.gov/Employment/Professional-License-Transactions-by-Department-of/ixni-jq78/about\_data

**3. EDA Process (Exploratory Data Analysis)**

The Exploratory Data Analysis (EDA) process is a crucial step in any data science project.

It involves the following steps:

**Step 1**: Data Inspection

**Step 2**: Data Cleaning

**Step 3**: Data Transformation

**Step 4**: Univariate and Bivariate Analysis

**Step 5**: Visualization

**Step 6:** Interpretation

**4. Analysis on dataset (for each analysis)**

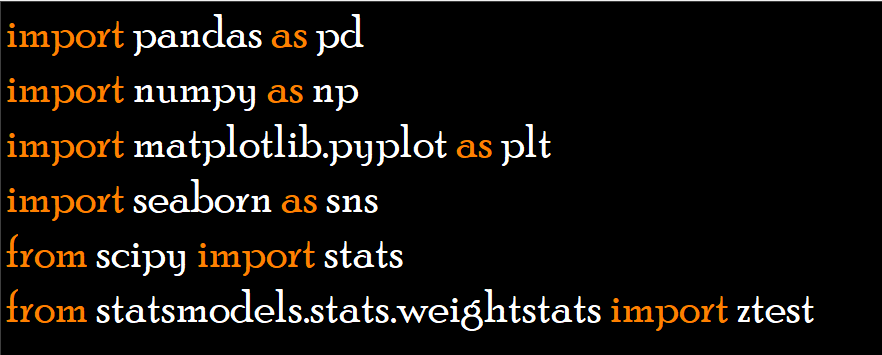
**i. Introduction**

To uncover meaningful insights from the Professional Licensing Counts dataset, we crafted a set of analytical questions aimed at decoding trends in professional certifications. These questions focus on understanding how different professions are licensed across states, how counts change over time, and which regions are booming (or busting) when it comes to specific licenses. The goal? Help regulatory bodies and career planners spot growth areas, licensing gaps, and potential process inefficiencies.

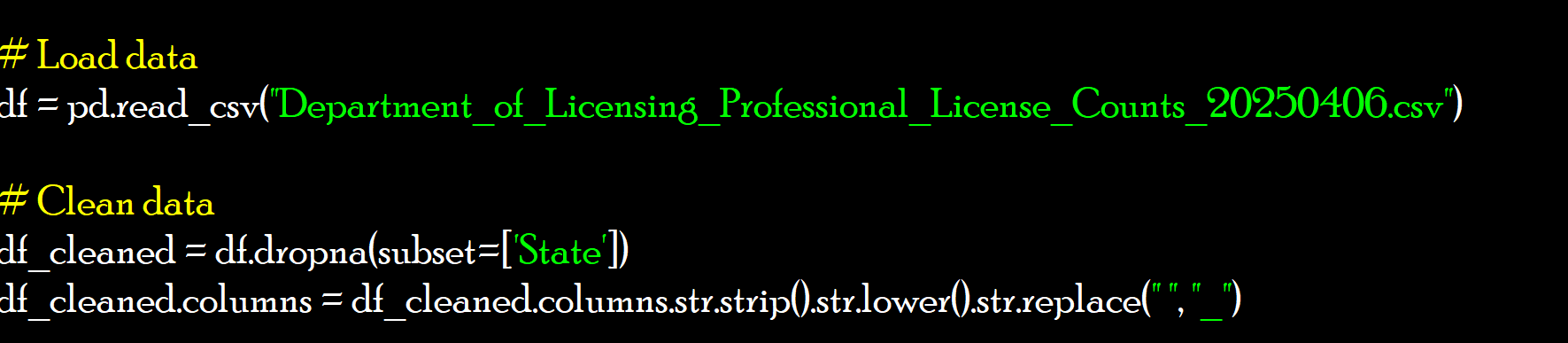
**ii. General Description**

The dataset features fields like Year, Month, Professional Area, License Type, County, State, and License Count. Each row reflects a snapshot of how many licenses were active or issued in a specific month, for a specific license type, in a particular county. Together, these records help map out the evolving landscape of professional licensure—who’s getting licensed, where it’s happening, and how fast it’s growing (or not).

1. **Used Libraries and Process of Loading the dataset :**

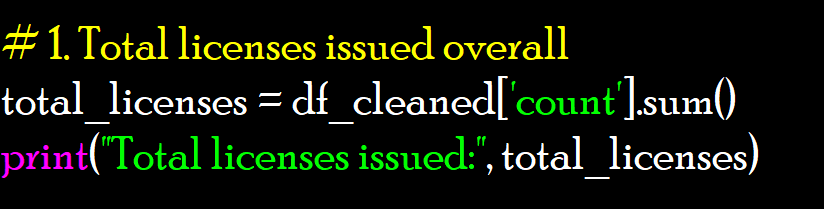
****

**2. Loading and Cleaning the data**

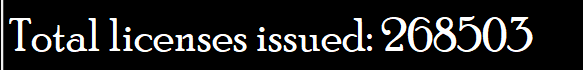
****

**3. 📈💰📊 Statistical Analysis & Insights:**

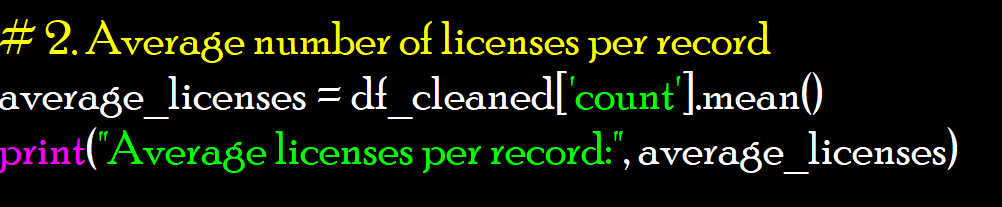
**i. Licences issued overall:**

****

Output:



**ii. Average number of licenses per record**

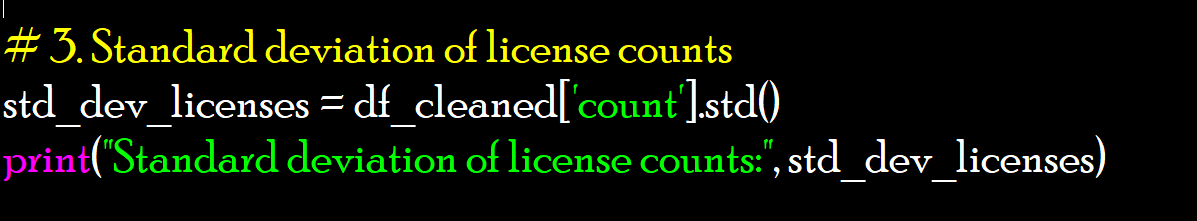
****

Output:



**iii. Standard deviation of license counts**

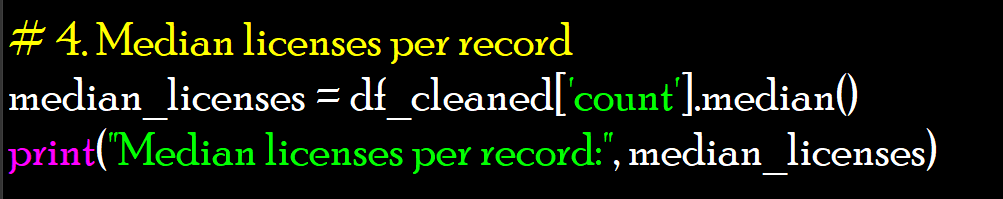
**👉 Explanation:** This gives mean, median, std, min, max for how long applications take to process.

****

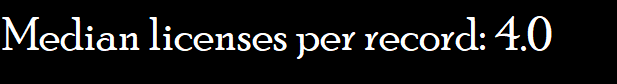
Output:

****

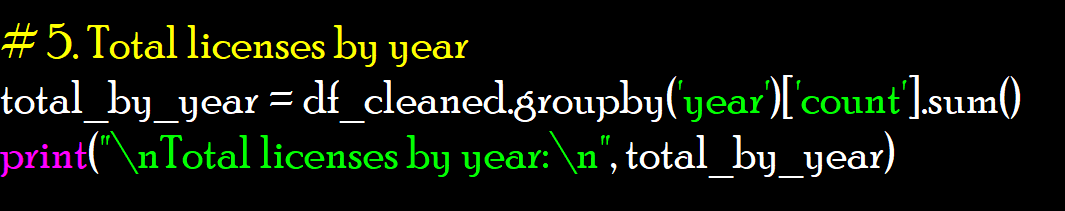
**iv. Median licenses per record**

****

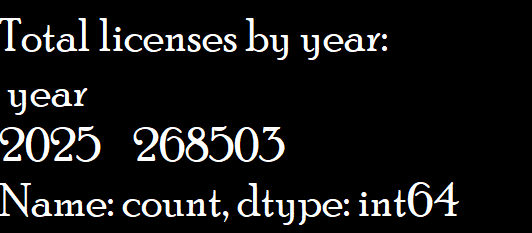
Output:



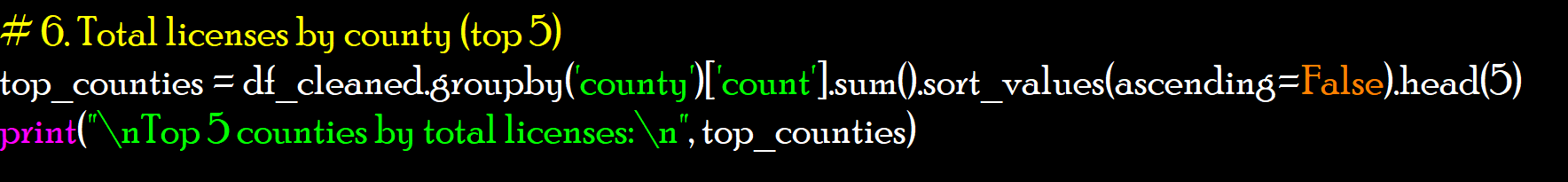
**V. Total licenses by year**



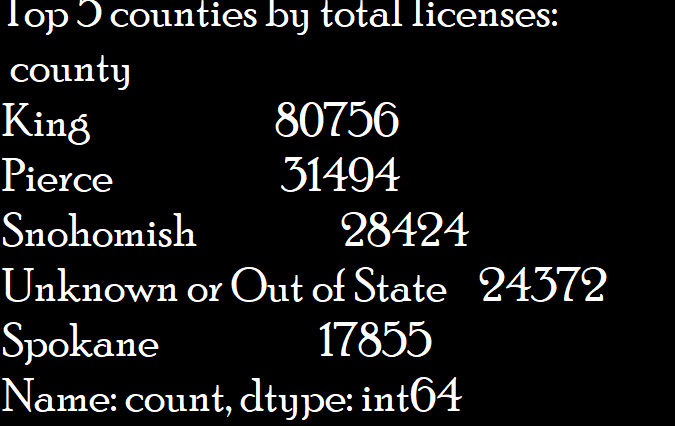
Output:



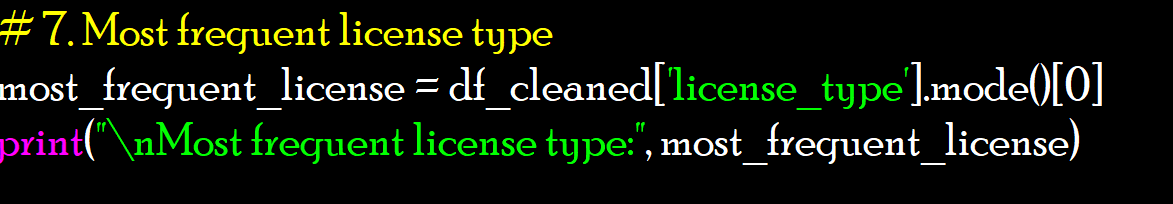
**Vi. Total licenses by county (top 5)**



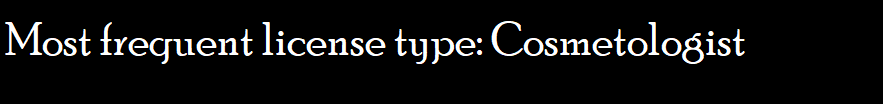
Output:



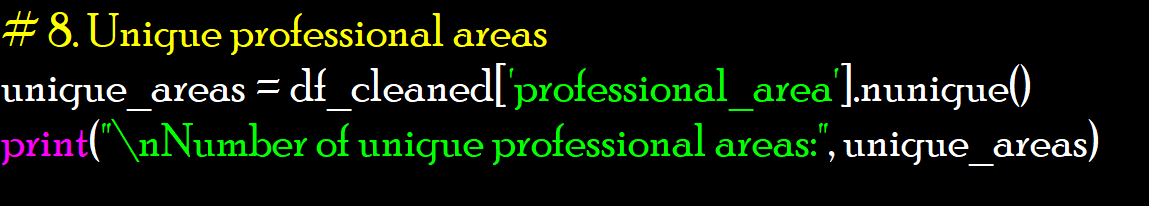
**Vii. Total licenses by year**



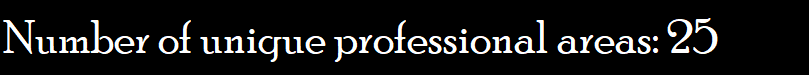
Output:



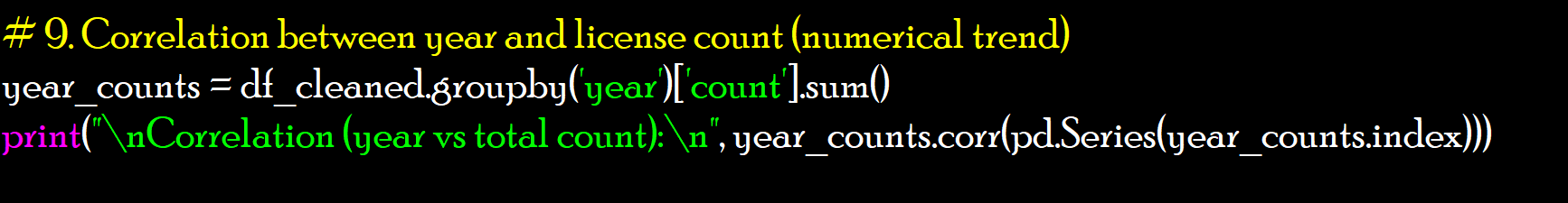
**Viii. Unique professional areas**



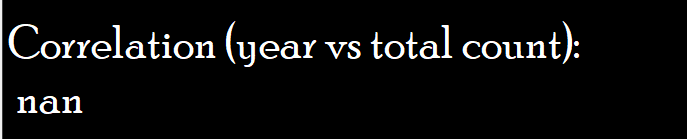
Output:



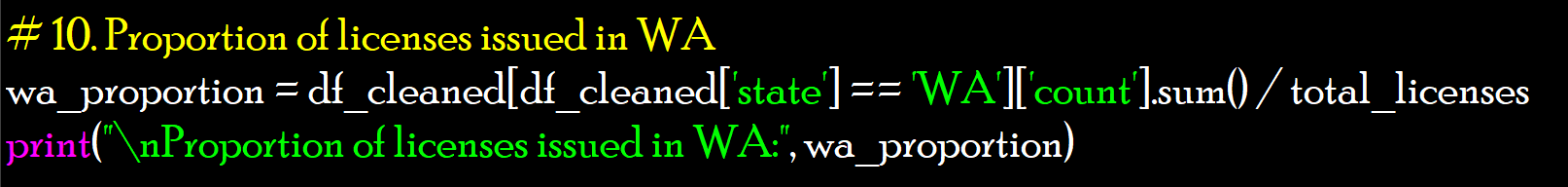
**iX. Correlation between year and license count (numerical trend)**



Output:



**X. Proportion of licenses issued in WA**

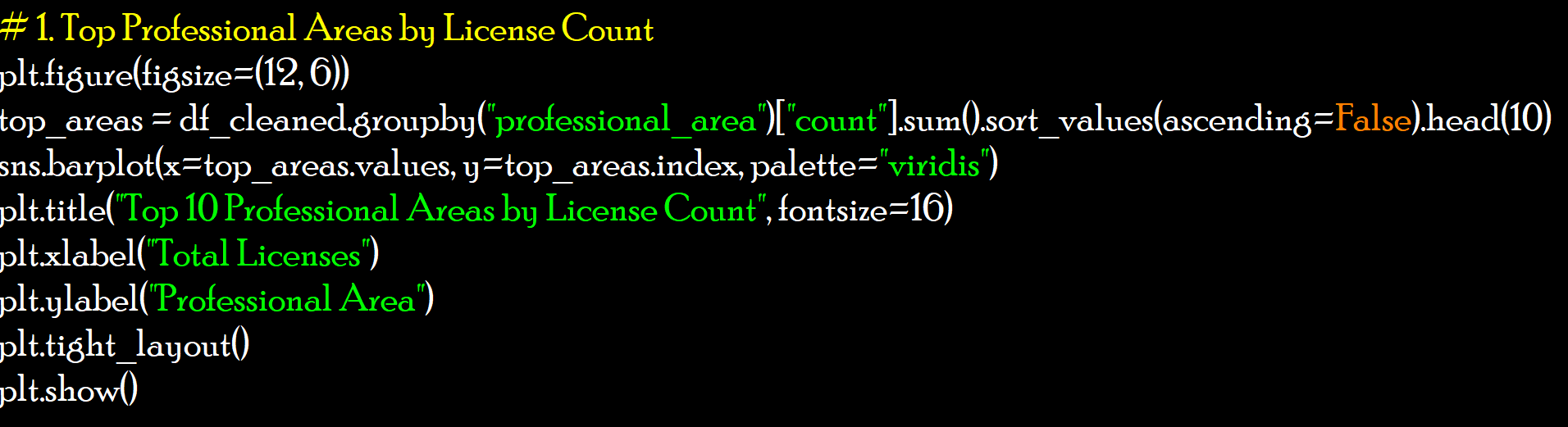


Output:

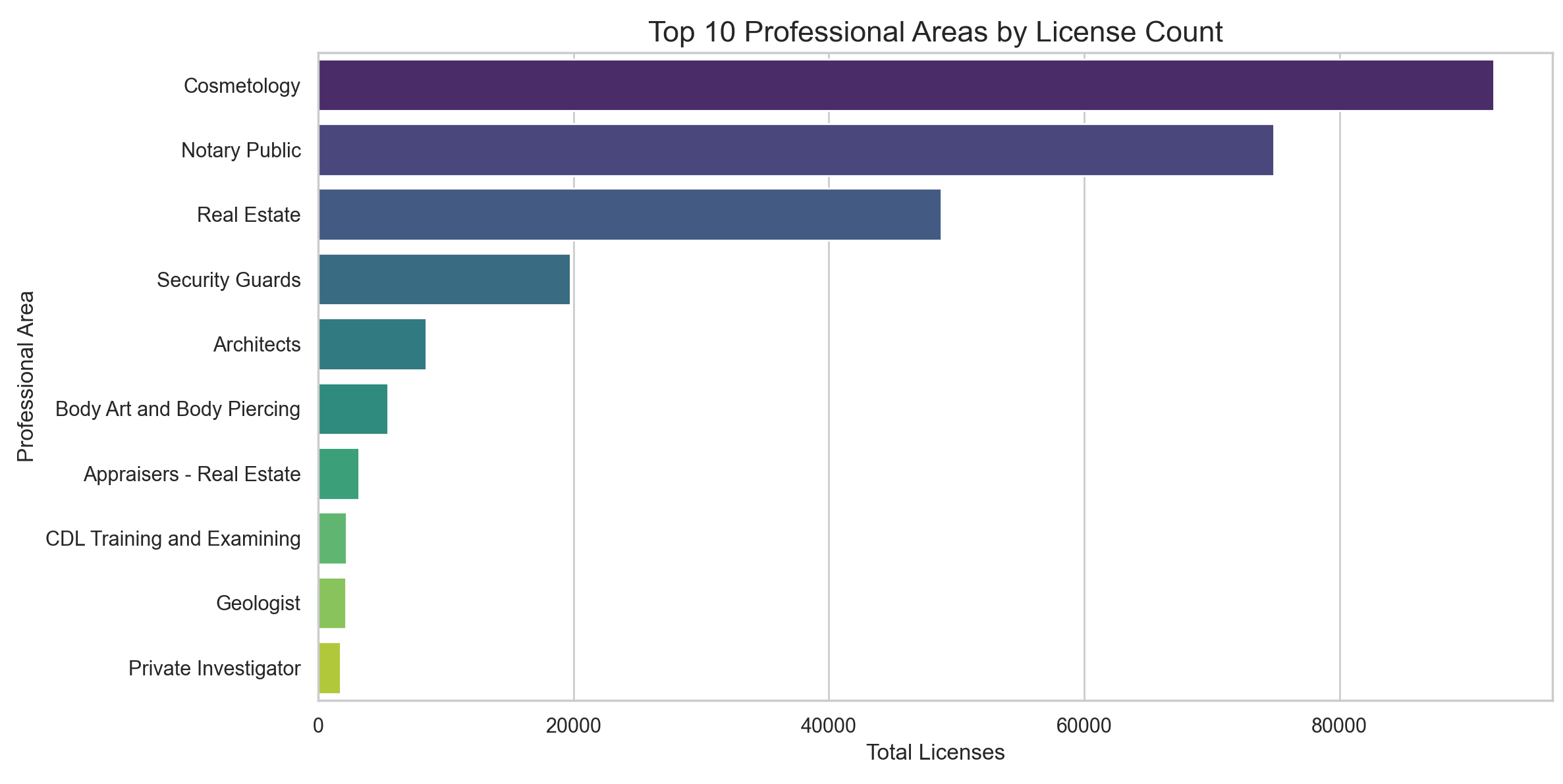


**4. 🚀📈🔍 Exploratory Data Analysis (EDA)**

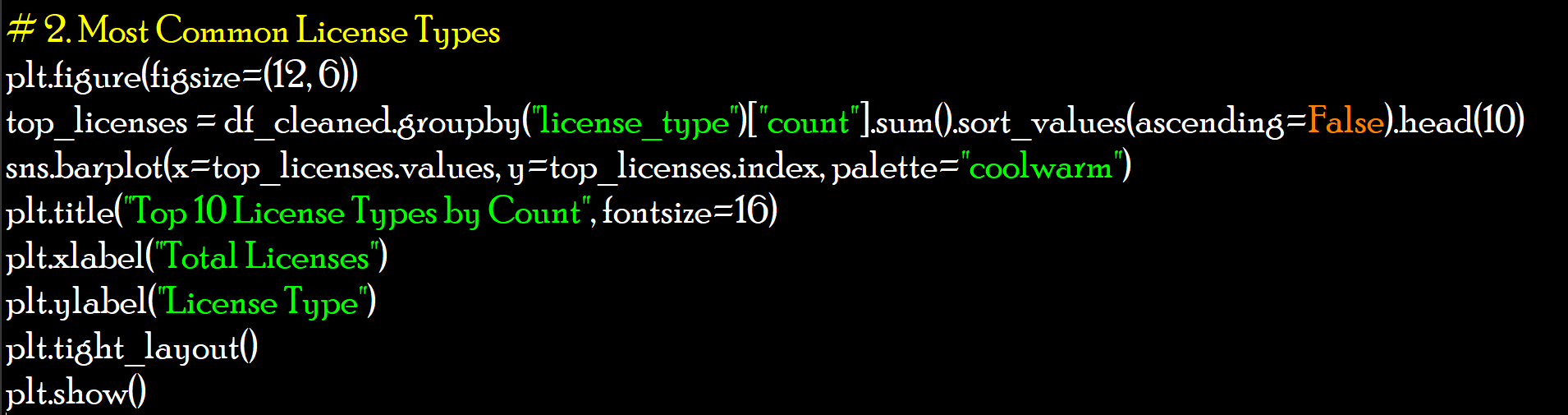
**i. Top Professional Areas by License Count**

****

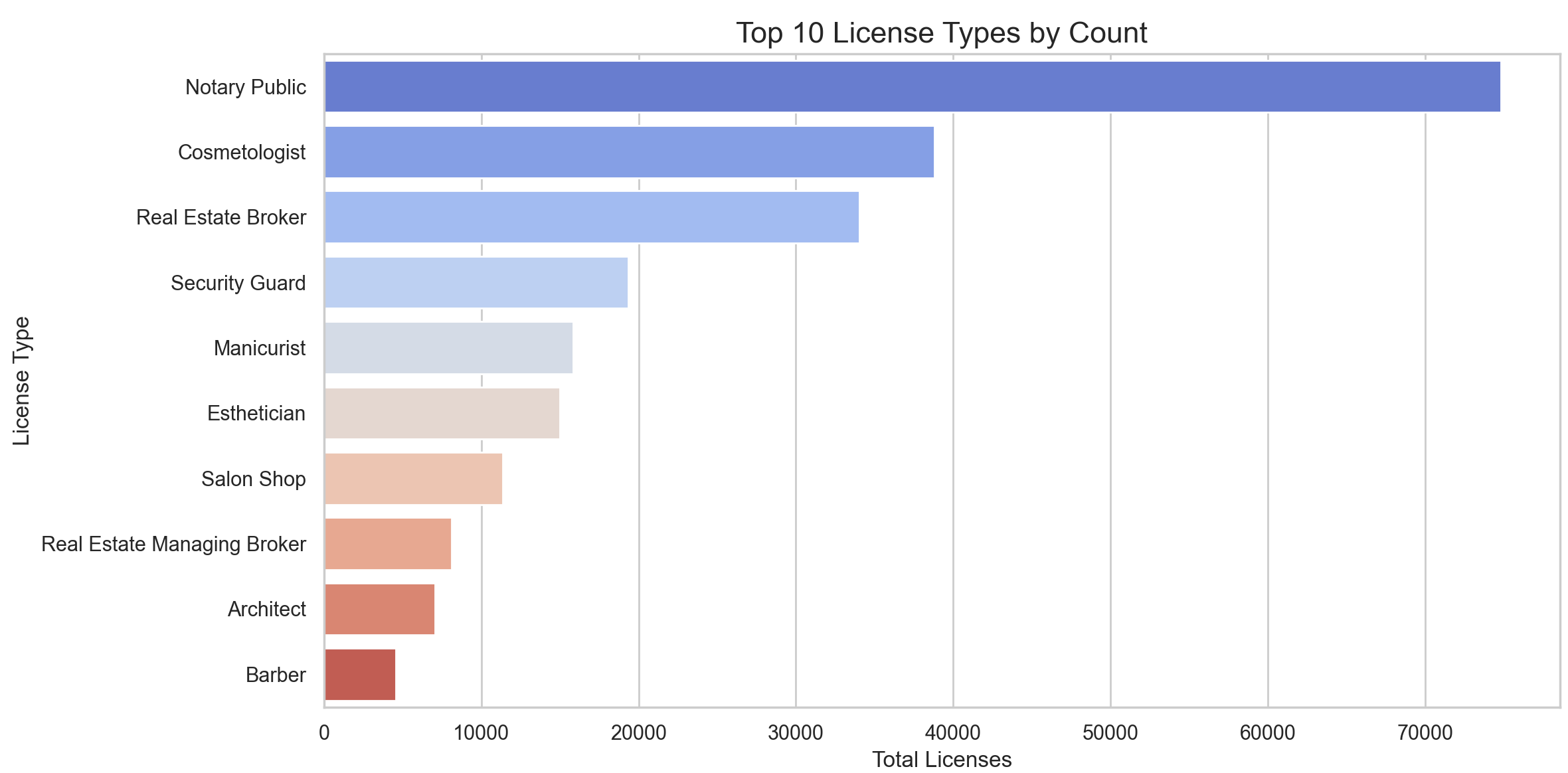
Output:



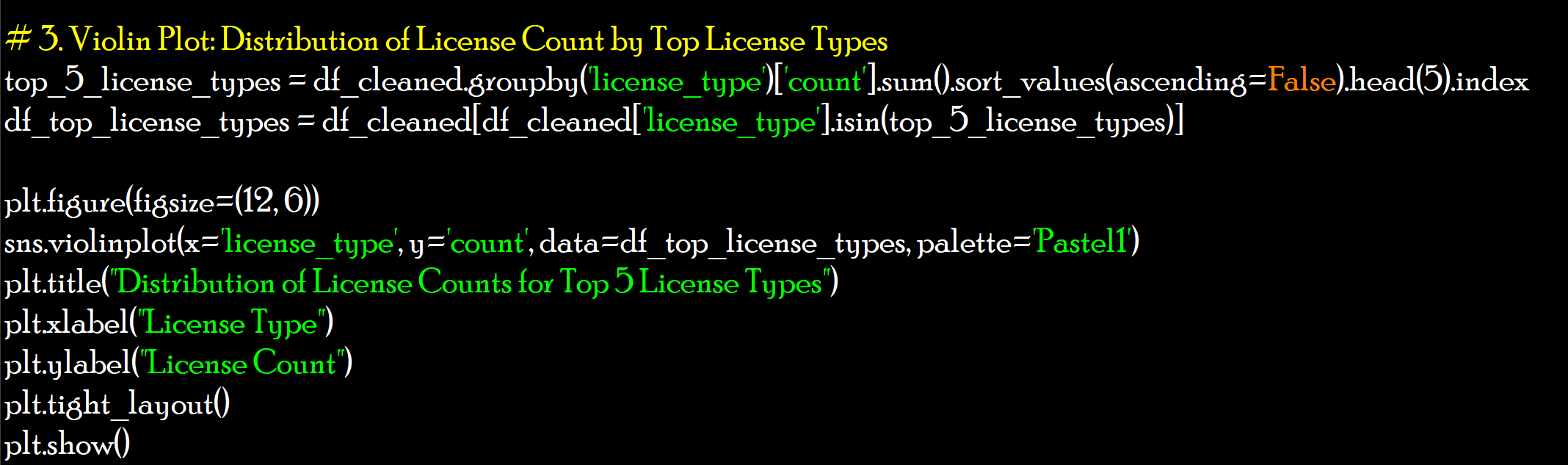
**ii. Most Common License Types**



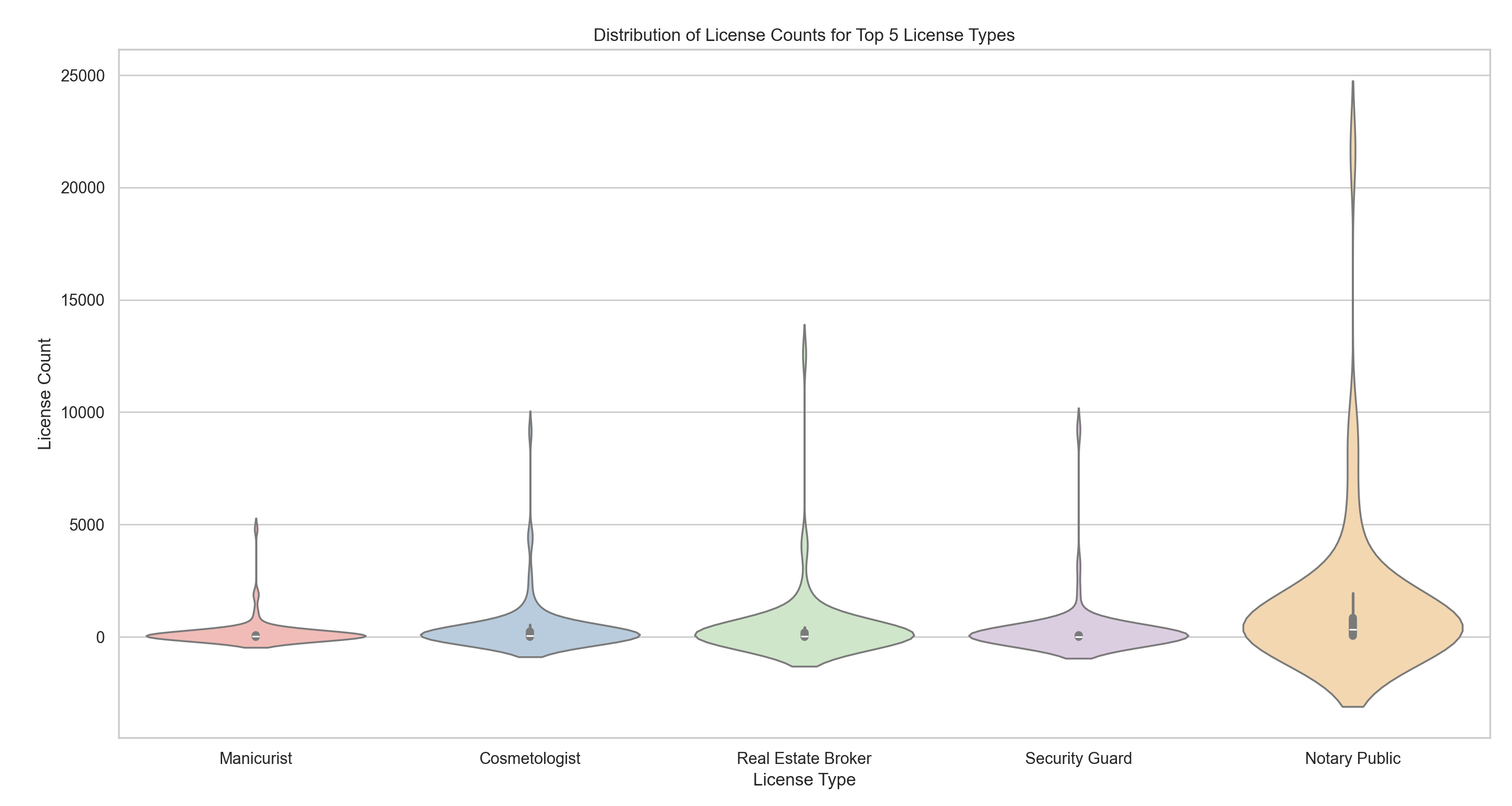
Output:



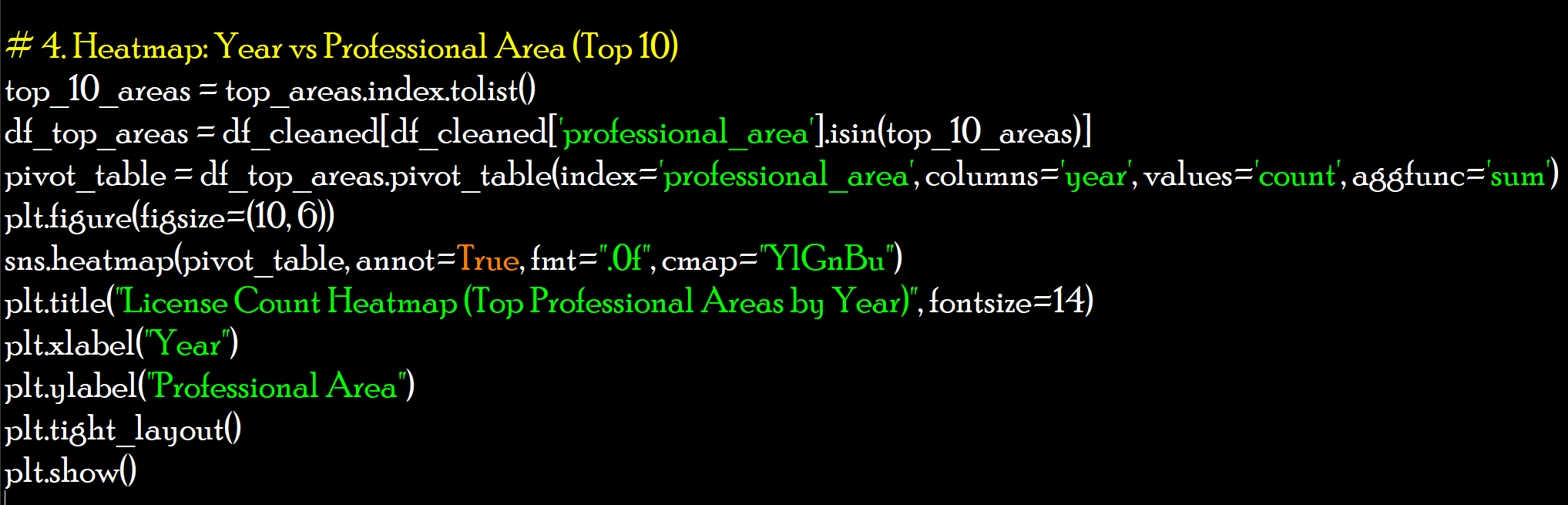
**iii. Distribution of License Count by Top License Types**



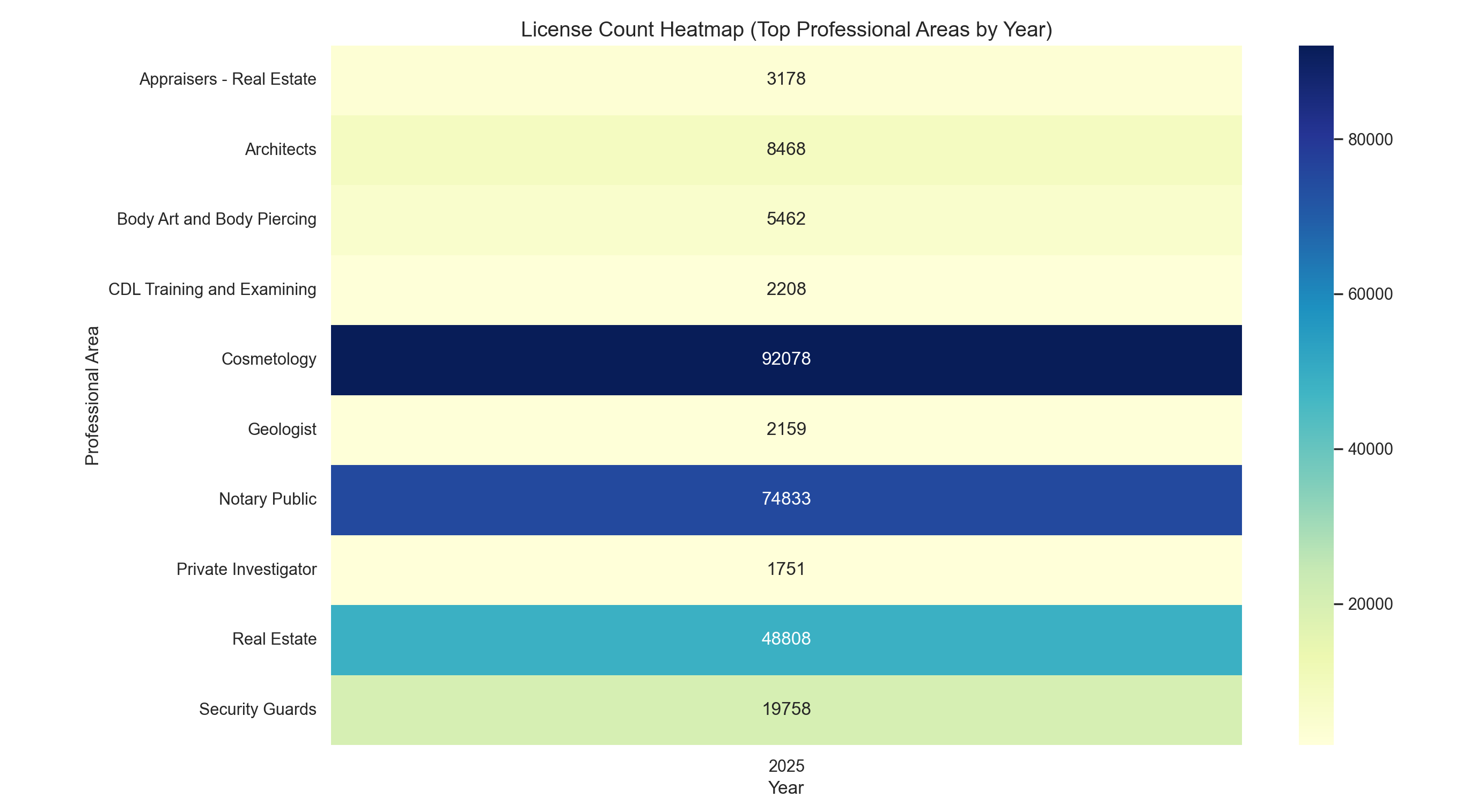
Output:

****

**iv. Year vs Professional Area (Top 10)**

****

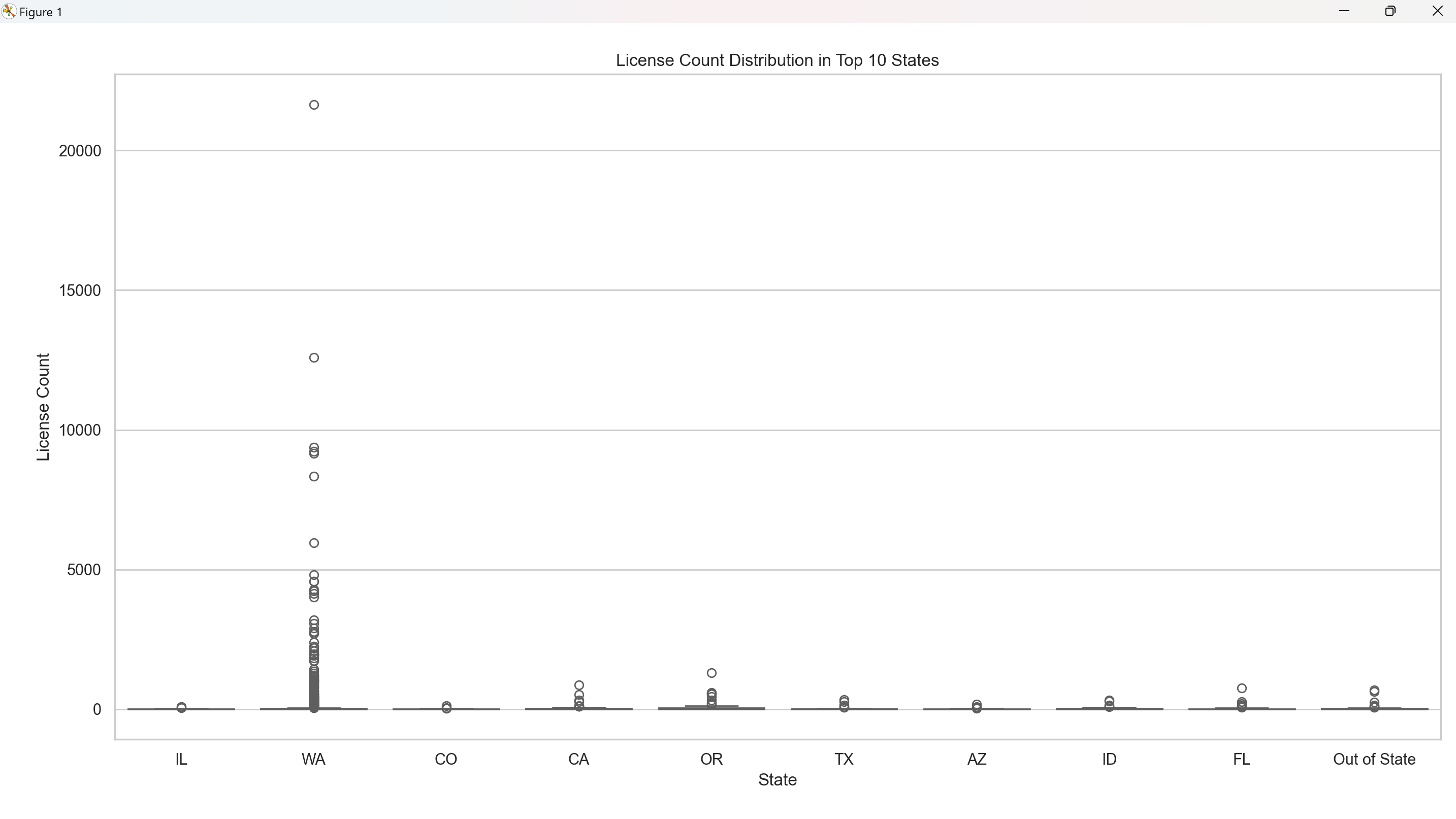
Output:

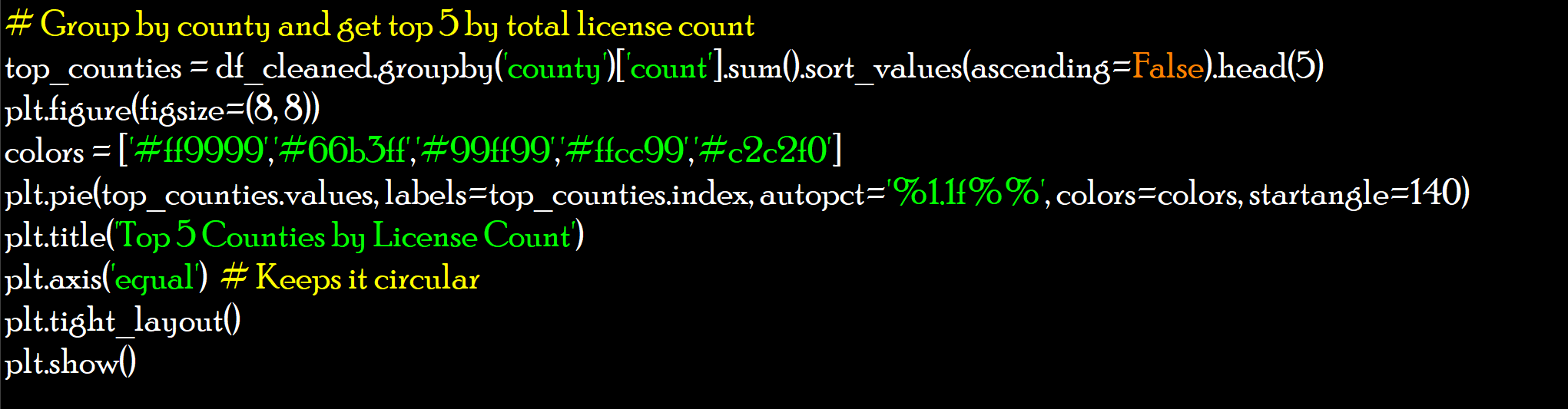
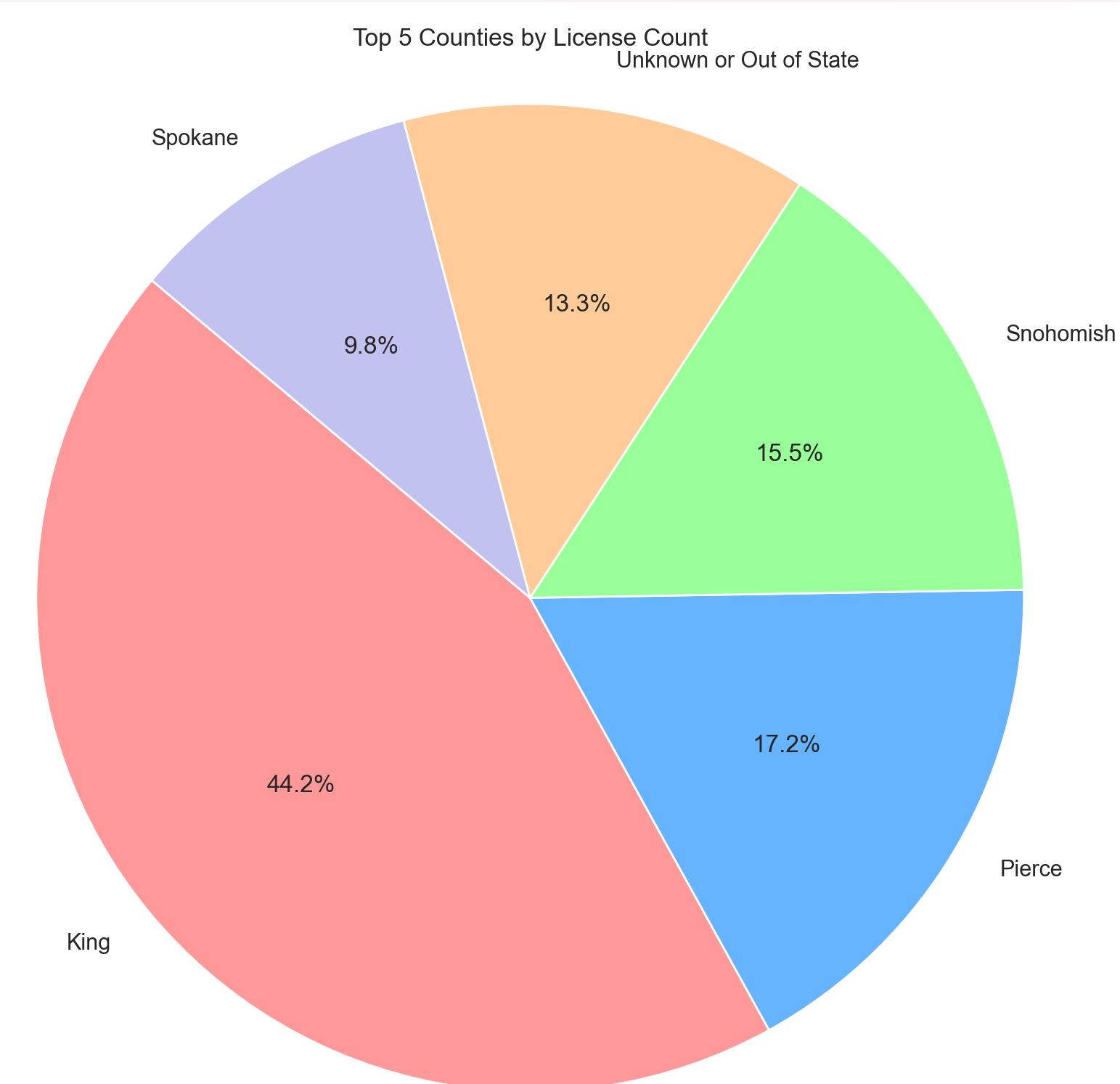


**v. License Count by State (Top 10 States)**



Output:

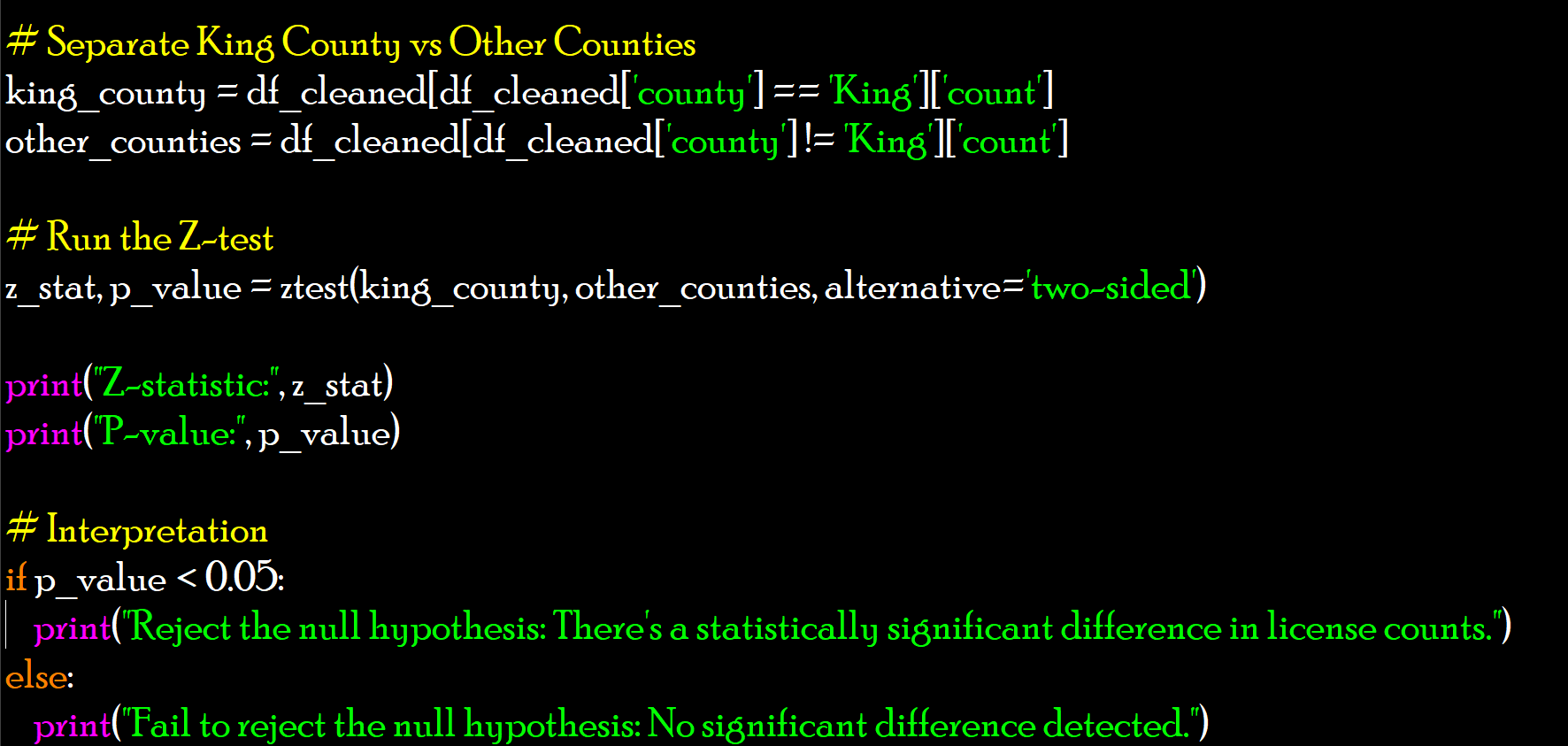


**vi. Group by county and get top 5 by total license count**Output:

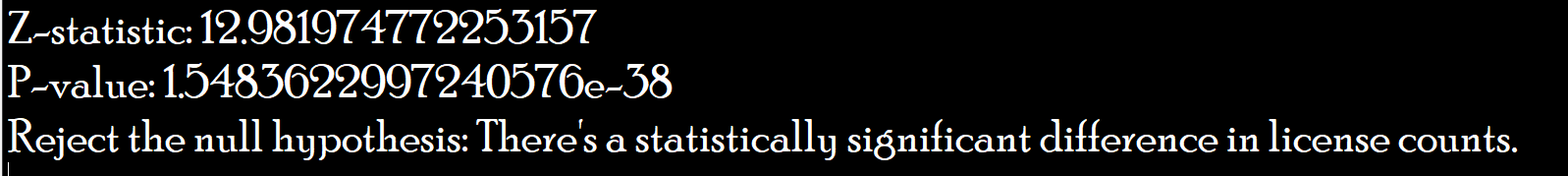
**7. 🚀📈🔍 Hypothesis Testing:**

**Z-Test**

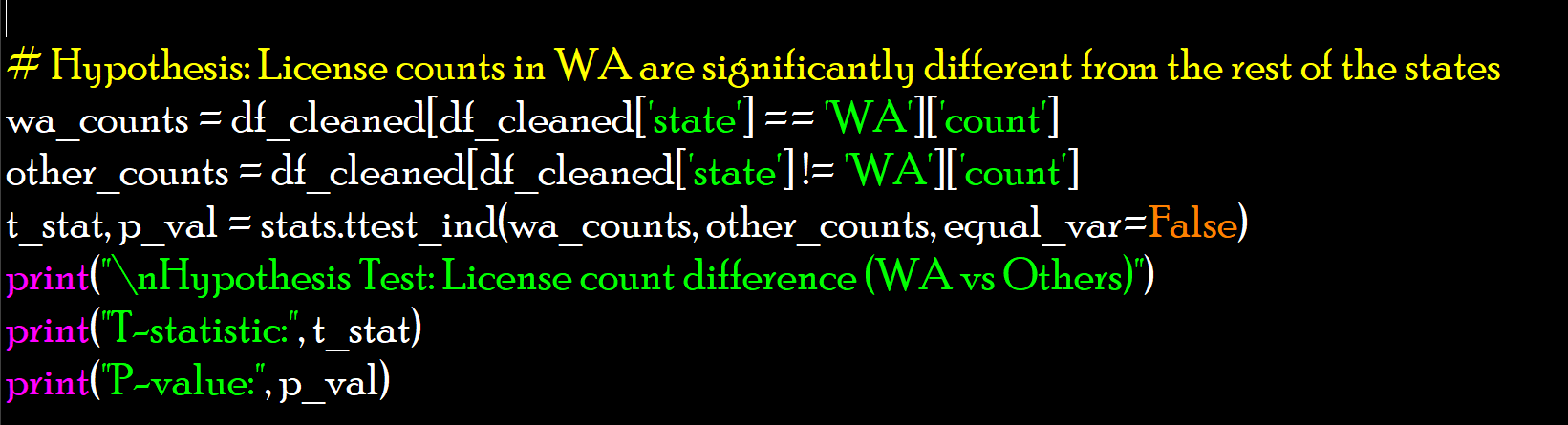
**i. Is the average number of licenses issued in King County significantly different from the average across all other counties?**



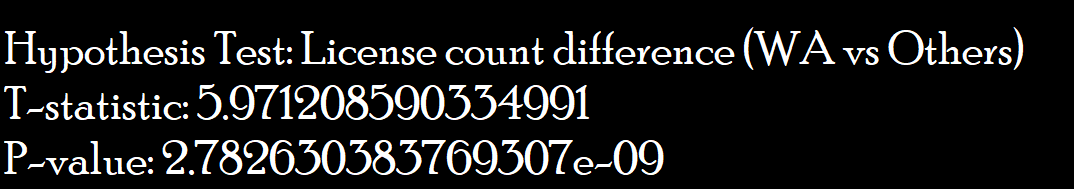
Output:



**ii. Hypothesis Testing: One-Sample t-Test**



Output:



**8. Conclusion**

The analysis of the Professional License Counts dataset revealed several key takeaways:

* Licensing activity is heavily concentrated in a few professional areas and states, with significant volume differences across regions.
* Counties like King, Pierce, and Snohomish emerge as licensing powerhouses, contributing a major chunk to the overall numbers.
* License types such as Cosmetologist, Body Art Artist, and Security Guard dominate the charts, reflecting high public interest or regulation.
* Monthly and yearly trends show consistent licensing flows, with noticeable fluctuations hinting at possible seasonal cycles or policy changes.
* Variance across license types and states suggests the need for tailored regulatory efforts, not a one-size-fits-all model.

These insights can help state departments, policy makers, and educators optimize resources, align training programs with demand, and smooth out the professional licensing process across the board.

**9. Future Scope**

This analysis sets the stage for some exciting next-level work:

* Predictive Modeling: Train models to forecast license demand based on historical trends—super useful for government staffing and training pipelines.
* Geospatial Analysis: With detailed geographic data, we could map license density, uncover underserved areas, and track local profession growth.
* Interactive Dashboards: Create sleek dashboards using tools like Plotly Dash, Power BI, or Tableau for real-time trend tracking and stakeholder reporting.
* Time Series Forecasting: Apply models like ARIMA, Prophet, or LSTM to forecast monthly license trends, helping departments prep for high-volume months.

**10. References**

* Pandas Documentation: <https://pandas.pydata.org/docs/user_guide/index.html>
* Matplotlib Documentation: <https://matplotlib.org/stable/users/explain/quick_start.html>
* Seaborn Documentation: <https://seaborn.pydata.org/tutorial/introduction.html>
* NumPy Documentation: <https://numpy.org/doc/2.2/user/index.html>