Iowa Liquor Sales Case Study

Data

Iowa publishes very detailed information about liquor sales within the state. You can find a CSV file of item-level sales by store for 2018-2021 here.

You can read more about the data on the Iowa Data Portal.

Questions

Download the Iowa liquor sales data for 2018-2021. You may use Python or R and any supplementary open-source packages of your choice. Clean, aggregate, and organize the data sufficiently to answer the following questions:

1. What impact did Covid have on the overall liquor market in Iowa?
    a. What trends evolved over the next 3-18 months?
    b. Was there a notable shift in the types of products purchased in terms of pack size
2. Which are the fastest growing types of liquor (e.g., vodka, tequila, rum, etc.)? How has market share changed over time?
    a. Write a function that takes a list of liquor types as an input and visualizes the market share over time for each of those.
    b. What is driving the growth in tequila sales? Increases in average price or increases in volume sold?
3. Grouping individual store brands together (e.g., all of Walmart, Liquor Barn, Hy-Vee, etc.), who are the top 10 retailers by year?
4. In late 2019 Heaven Hill Brands bought a portfolio of liquor brands from Constellation Brands. What percentage of Heaven Hill's growth in 2020 can be attributed to the acquisition?
5. What data integrity issues did you discover? How could you (or how did you) solve/account for these?
    a. Comment on any data issues you discovered and what assumptions you used to deal with them.
    b. What's a simple solution to solving data quality issues?
    c. What would be a more scalable solution to deal with quality issues?

Deliverables

Please share a reproducible notebook or file(s) along with the version of packages used to answer the above. Also include a PDF or Word doc outlining your answers to the above question (no more than 2 pages, please).

In addition to the write-up, files can include Python/R files, a Jupyter notebook, Hex project (or similar hosted notebook), etc.