# Diageo Business Analytics Case Study



Diageo PLC is a British multinational alcoholic beverages company known for its vast portfolio of globally renowned spirits and beer brands. In your role as a business analyst, you have been presented with a project of immense practical value - to conduct an Exploratory Data Analysis (EDA) on the corporation's liquor transactions and competitor transactions within the state of Iowa. You are expected to perform robust EDA over time to identify trends, specific items, locations and other novel insights that would aid strategic business leaders at Diageo. Your audience includes analytics leadership, and sales leadership with less data-centric acumen. As a result, your findings need to cover both analytical efforts and their business impact for the personnel selling in the region.

## Your Objective

Your key objective is to identify notable sales trends, key customer segments, and high-performing locations that could assist corporate planners in strategizing production, sales incentives, and various other business elements. You can also understand trends, the impact of the pandemic, changes in prices, popular brands, item aspects and other relevant business aspects of the data that would help with business planning for the sales and marketing team, the production and operations team and even the logistics and supply chain personnel.

## Data

Source: In Iowa liquor sales is a state regulated transaction. As a result all invoices are freely available from the state’s website: <https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy>

This case is a subset of the 27M transactions dating back to 2012. The case data is limited to 2019-2023 and segmented by month, less December 2023 which was unavailable while the case was being constructed. It is good for EDA due to its diversity of categorical and numerical columns like item, category, vendor, and more while also having a temporal component.

The case data folder has subdirectories 2019, 2020, 2021, 2022, 2023. Within each subdirectory, there are 12 files, labeled such as ` 1\_Jan\_2019\_Iowa\_liquor\_sales.csv`.

In total there are more than 12M records. As a result, some students may struggle with how to perform EDA on a laptop efficiently as computation may be limited by RAM. Students will be asked to think creatively about performing EDA with this realistic yet challenging data.

The column names are not easy to work within R due to spaces and parenthesis. As a result, students are encouraged to use the function `make.names()` in their workflow.

## Data Dictionary

To aid in your analysis, you have access to a comprehensive dataset containing the following variables:

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Example |
| Invoice/Item Number | Unique identifier | INV-16685400042 |
| Date | Month/Day/Year date | 2019-01-02 |
| Store Number | A numeric identifier of a store | Store Number |
| Store Name | A unique text-based store identifier | HY-VEE FOOD STORE / DUBUQUE |
| Address | The address of the purchasing store | 3500 DODGE ST |
| City | An Iowa city categorical variable | DUBUQUE |
| Zip Code | The USPS mailing zip code where the purchasing store is located | 52001 |
| County | Iowa has 99 counties representing geographic locations. This variable holds the name of the county where the sale is occurring. | DUBUQUE |
| Category Name | Product category of the liquor purchased | AMERICAN VODKAS |
| Vendor Name | The name of the liquor distributor executing the sale | SAZERAC COMPANY INC |
| Item Description | Specific item purchased | TITOS HANDMADE VODKA |
| Pack | Numeric representing the number of bottles in a item’s case | 12 |
| Bottle Volume (ml) | Individual bottle volume measured in milliliters | 1000 |
| State Bottle Cost | The state regulated purchase cost the distributor can charge, keep in mind this may change over time. | 12.67 |
| State Bottle Retail | The state regulated store price of a bottle. This represents the maximum a store can charge a consumer. | 19.01 |
| Bottles Sold | The total number of bottles purchased on the specific invoice. | 36 |
| Sale (Dollars) | The total sale for the invoice. Thus this is the `State Bottle Cost` times `Bottles Sold` | 684.36 |
| Month | An integer 1-12 representing the month associated with the sale. It has been extracted from the `Date` variable and used in segmenting the files. | 1 |

## The Task at Hand

Through careful EDA, you are entrusted with not just identifying trends, but also revealing insightful information that could reshape the company's understanding of its best performing products, preferred consumer choice, and geographical strengths within the state.

Your task is to delve into the transactional data and unravelbusiness relevant insights **(though not limited to these suggestions):**

1. Sales trends across time - seasons, months, weeks, weekends, etc.

2. Dominant store locations and their performance characteristics.

3. Prominent liquor categories and their sales composition.

4. Consumer preferences in terms of product features such as bottle volume, price segments, etc.

5. Vendor performance and their contribution to total sales and liters.

6. Impact of geographical and other factors on sales.

## Analytical Implementation

After thorough examination of the dataset, you will need to highlight significant patterns and insights. Proceed to frame a narrative describing how these insights could translate into tangible business strategies.

Remember, your mission doesn't end with just identifying patterns. Strive to dig deeper and uncover the 'why' behind these trends, in ideal situations supported with external data sources such as retail sales trends, or cited economic factors that support your findings. How does consumer behavior change in different regions or time periods? Which products are consistently high-performing and why? While you may not have conclusive evidence, you’re asked to support your findings with data, and have some supporting information as a possible and plausible explanation such as overall liquor sales, socio-economic factors in some communities, or the impact of externalities like the pandemic.

## Presentation of Findings

Prepare a convincing PowerPoint presentation to share your findings and insights with the corporate planning team. This should include your observations, supporting data visualizations, and potential implications for business strategies.

To ensure your analysis is comprehensive and robust, it will be subjected to a validation process which could incorporate comparisons to actual sales results, in-field testing, or reviews by other experts within the organization. Your findings could be the key to informing major business decisions at Diageo, highlighting the importance of rigorous analysis and logical reasoning. Thus, the audience is wide and you will be expected to support your information with data, be consultative to the business and project confidence in your analysis and findings.

You presentation should be no more than 10min in length.

## Final Note

Your insights will act as a robust foundation upon which other departments in the organization can strategize and make informed decisions. Let the data lead the conversation and provide an unbiased perspective on the opportunities and challenges that lie before Diageo.

## Criteria for Success

|  |
| --- |
| Organization – Was the presentation well organized? |
| Delivery – Was the content delivered clearly and persuasively with the audience in mind? |
| Code Documentation – Was the data mined to support the conclusion? |
| Written Supplemental – Is it grammatically acceptable, organized and error free.  -Is the data supported clearly and coincides with the data, and narration while being contextualized with external information? |
| Data Mining Process – Overall, as a complete portfolio of work, is the topic interesting, organized, researched, supported and delivered effectively? Was CRISP-DM, SEMMA, or a similar workflow followed to organize the work if appropriate? |

**Delivery and Narration Guidance**

You are not allowed to use an ai avatar, or speech creation for narration. While this is certainly useful technology in many instances, the purpose of the business case presentation is to improve *your* presentation skills. In a business setting you will still be expected to articulate your findings and not send an avatar for this type of business meeting. As a result services like <https://elevenlabs.io/> or <https://www.heygen.com/> are not permitted. Your “boss” or “audience” in the case will not accept these technologies.

**Written Supplemental Guidance**

Submit a document to represent the entirety of your presentation including the data, process, findings, and implications in a business setting. Thus it’s a professional report, anything less than a professionally written and organized report will be considered sub-optimal. Amazon for example doesn’t use PowerPoint and instead uses “6 pagers” to make business recommendations, as such some organizations prefer written information over presentations. The use of external and verifiable sources is expected to add context and support any component of the paper. **The minimum is 2 pages maximum is 5**. **Double spaced and 12 point font.**

Helpful tips:

Markdown is not encouraged. It is to be a professional report similar to amazon's 6 pagers though does not have to be as long. As a business report consumed throughout many departments, code and screenshots are discouraged/to be limited and instead describe your intentions/problem statement, data aspects, and results/findings/implications in an organized manner. Lastly, outlines and bullet points alone will not earn you robust marks.

Overall guidance:

You are expected to submit R code, a voiceover narration of a live business presentation (can be a standalone video file, or inserted within the powerpoint) , a set of slides, and a written supplemental. If you submit documents with links to youtube videos or cloud drive files, you must ensure all links are accessible. Links set to private which are not viewable or downloadable will automatically result in 0 for that section of the rubric.