Market Segmentation Using K-Means Clustering Analysis

Cincinnati Raptors

Sudhir Vedula
Romil Shah
Hardik Patel
Nikit Choksi
Shiv Patel

Use Case

What specific business problem is this software trying to solve in the capacity for which you are using it?

The ABC Beer Company is trying to increase its revenues. The company has identified that to improve its revenue, they should be increasing their sales. The best way to increase sales is by analyzing the market segmentation using K-means clustering. To perform this analysis, we can do it by using Alteryx Designer in collaboration with Tableau Public.

Who are / would be its users?

Alteryx is mainly used by business users and analysts of an organization to answer business questions that are time-consuming, manual, error-prone and risky.

Tableau Public is a free platform from the Tableau Product Suite. It being a visualization tool is used by companies to create interactive dashboards with the data. It helps in improving the efficiency.

Why is the problem relevant for businesses?

For a company to sustain and grow, it is quite evident that it is business should generate more revenues either by increasing their sales or improving their services. The improvement in sales and services has direct impact on bigger profits. The ABC Beer Company is also, one of those companies who wants to improve their sales revenues to be in the Beer Business and sustain from the competition.

Why would a business want to spend resources on the software given other conflicting priorities?

Alteryx is ideal for users that want quick access to data to obtain results in a more consistent and timely manner. Executives, managers, employees, other departments, vendors, consumers, government, regulators, auditors, and others ask questions of marketing, sales, accounting, and HR, among other departments. Getting data together for individuals to make decisions is likely costing your business hundreds or thousands of hours each year. Anyone who spends time dumping reports into Excel for manipulation or getting data from databases may save enough time using Alteryx to warrant the cost.

Tableau public is a free tool, there is no cost involved. It allows to create interactive data visualizations for the web integrated with spreadsheet or file as data source. These visualizations can be shared with various users.

What data are you using and what is it source?

To perform Market Segmentation using predictive analytics and K-Means clustering, based on the qualitative values IBU(International Bitterness Units) and ABV(Alcohol By Volume) of each beer, we have used Excel files provided by OpenBeerDB.com. Following is the basic structure of the data:

Beer.xlsx

 ➢ Records: 5901
 ➢ Columns:

Id	brewery_id	name	cat_id	style_id	abv	ibu	srm	upc	filepath	add_user	last_mod

Breweries.xlsx

 ➢ Records: 1414
 ➢ Columns:

Id	name	address1	address2	city	state	code	country	phone	website	filepath	add_user	last_mod

styles.xlsx

 ➢ Records: 141
 ➢ Columns:

id	cat_id	style_name	last_mod

geocodes.xlsx

 ➢ Records: 1304
 ➢ Columns:

id	brewery_id	latitude	longitude	accuracy

Software Details

What are the core (IT) components/features of the software?

Main features of Alteryx are:

- ➢ Data Analysis
- ➢ Data Cleansing
- ➢ Data Connectors
- ➢ Data enrichment and insights
- ➢ Data Joining
- ➢ Data Mining
- ➢ Data Profiling
- ➢ Data science and decisions
- ➢ Data Source Connectors
- ➢ Data Source Integrations
- ➢ Ease of Integrations
- ➢ ETL Capability
- ➢ ETL Scheduler
- ➢ In-memory data model
- ➢ Machine Learning
- ➢ ML Algorithm Library
- ➢ Mobile Reporting
- ➢ No-Code Sandbox
- ➢ Publish to PDF
- ➢ Self Service Analytics
- ➢ Self Service Data Preparation
- ➢ Self-service Analytics
- ➢ Visual Workflow Management

Source:

https://downloads.alteryx.com/Documentation/Alteryx%20Technical%20Overview.pdf

Main features of Tableau are:

➢ Tableau Dashboard

Tableau Dashboards provide a wholesome view of your data by the means of visualizations, visual objects, text, etc.

➢ Collaboration and Sharing

Tableau provides convenient options to collaborate with other users and instantly share data in the form of visualizations, sheets, dashboards, etc. in real-time.

➢ Live and In-memory Data

Tableau ensures connectivity to both live data extraction from external data sources as in-memory data. Tableau provides additional features to support data connectivity such as automatic extract refreshes, notifying the user upon a live connection fail, etc.

➢ Data Sources in Tableau

Tableau offers a myriad of data source options you can connect to and fetch data from. Data sources ranging from on-premise files, spreadsheets, relational databases, non-relational databases, data warehouses, big data, to on-cloud data are all available on Tableau.

➢ Advanced Visualizations (Chart Types)

One of the key features of Tableau and the one that got its popularity is its wide range of visualizations. In Tableau, one can make visualizations as:

- Bar chart
- Pie chart
- Histogram
- Gantt chart
- Bullet chart
- Motion chart
- Treemap
- Boxplot

➢ Maps

Tableau has a lot of pre-installed information on maps such as cities, postal codes, administrative boundaries, etc.

➢ Robust Security

Tableau takes special care of data and user security. It has a fool-proof security system based on authentication and permission systems for data connections and user access. Source:

https://data-flair.training/blogs/tableau-features/

How does the software work?

Alteryx Designer:

Alteryx Designer is a Windows software application that provides an intuitive drag-and-drop user interface for users to create repeatable workflow processes. Users can drag tools from a toolbox onto a canvas, connect them together, and edit their properties to create Alteryx workflows, apps, and macros.

The unique drag-and-drop workspace within Alteryx allows users to run workflows as they are being developed to instantly access and process data. This allows for continuous enhancements and refinements to the workflow and data results. Additionally, users can view data anywhere within a created workflow stream as a table, map, graph, or other visual representation. Alteryx Designer is written mostly in C# and executes the workflows through a local instance of the Alteryx Engine. In an Alteryx Server deployment, the Scheduler interface component within Alteryx Designer allows users to schedule workflows to be executed at predetermined times or specific recurring intervals.

The Scheduler then communicates with the Alteryx Service where jobs are queued for execution at the appropriate time. Additionally, users may use Alteryx Designer to publish their workflows, apps, and macros to the Alteryx Analytics Gallery where others can have access to running or downloading them.

Source: https://downloads.alteryx.com/Documentation/Alteryx%20Technical%20Overview.pdf

Tableau*:*

Tableau connects to data sources, extract data into sources and make visualization of the data.

There are two types of data extraction-

1) Live

2) Extract

Live data connectivity is about extract data lively (online), an analyst will work on data and share a dashboard with the user. Users can read the data using the tableau reader.

Data extraction can be done from Tableau Desktop (offline) and it is published on Tableau Server. Users can access data using the Tableau server from any location.

Source:

https://www.educba.com/how-to-use-tableau/

How does the software integrate into an existing BI&A IT architecture?

Alteryx Designer:

Alteryx can be expanded further and integrated into other systems by way of Email and Run Command tools and events.

Email

Using the Email tool or event, Alteryx can prepare outgoing emails and submit the emails through the SMTP protocol to any SMTP-enabled email server.

The user can:

• Set the To, From, CC, and BCC fields.

• Create the subject line.

• Create the body of the message.

• Attach documents.

Run Command

Using the Run Command tool or event, Alteryx allows for the execution of command line functions through the "Run" shell. Common uses for this command include integration through:

• Batch commands to execute external processes.

• Complex file transfers.

• Custom scripts or other add-ons.

Source:

https://downloads.alteryx.com/Documentation/Alteryx%20Technical%20Overview.pdf

Tableau

Tableau helps users bring more powerful analysis with extensibility and embedding capabilities.

One can integrate applications into Tableau Dashboards using the following:

➢ Data Science Integrations
➢ Dashboard Application Integrations
➢ Embedded Analytics

It provides integration through data storage and processing, data prep and transformation, cataloguing and enterprise metadata management, data science, query acceleration.

Source:

https://www.tableau.com/products/our-integrations

Key alternatives to the selected software

How would you describe the broad category of your use of the software e.g., data visualization?

Alteryx Designer software can be categorized as one of the leading solutions for data preparation, data blending and advanced analytics.

Tableau Public can be categorized as an interactive data visualization software, focused on business intelligence.

Who are the market leaders for the broad category?

Market Leaders in the broad category of data preparation, data blending and advanced analytics are SAS, SAP Advanced Analytics, Salesforce, Microsoft, IBM, Oracle, Teradata and Informatica.

Reference:

https://www.appsruntheworld.com/top-10-analytics-and-bi-software-vendors-and-market-forecast/

Market Leaders in the broad category of data visualization are Microsoft Power BI, Qlikview, SAP Advanced Analytics & Tableau.

What are the key differentiators of the software/vendors in your software category?

Some of the key differentiating factors of Alteryx Designer are as follows:

1. Automated data preparation, blending, and analytics.
2. Drag-and-drop analytics for Snowflake, Excel, Databases, and more databases.
3. Data Science functionality without needing to know the code.
4. Customizing analytics with Alteryx SDKs.
5. Creating visual analytic workflows of any business process.

Some of the key differentiating factors of Tableau Public are as follows:

1. It allows you to create custom maps, charts, and graphs in clicks.
2. It allows you to explore visualizations to see how they were made. Tableau Public dashboards are visible to public and can be downloaded or browsed for reference.
3. Option to practice and improve skills with peer learning activities.

Software recommendation

What are the key conclusions you would provide about the software if asked by your supervisor or colleagues?

Alteryx Designer allows you to simply construct repeatable processes using a user interface. Alteryx Designer is simple to use. It can connect to numerous systems, cleanse, integrate, and generate analytics findings.

- ➢ **Input/Output**: Access data and deliver it in various formats from a variety of locations, such as a local desktop.
- ➢ **Join**: Regardless of data structures and formats, combine multiple data sources using a variety of tools.
- ➢ **Preparation**: Drag and drop tools make it simple to prepare data for further processing or analysis.
- ➢ **Spatial**: Provides a large amount of information to help you understand and make decisions
- ➢ **Predictive**: Provide access to over thirty pre-packaged prediction and forecasting tools.
- ➢ **Investigation**: Before beginning an analysis, it is helpful for analysts to understand the details of the data.
- ➢ **Parse & Transform**: Assists in reorganizing and formatting of the data for further analysis.

Tableau is simple to use, and because of its advanced features, creating a dashboard on it is more feasible and flexible.

- ➢ **Usability:** It is simple to access from a variety of sources, there is no need for technical or programming knowledge, and it is free.
- ➢ **Accessibility:** The data file can be downloaded locally on mobile or desktop for easy access and analysis, multilingual data representation, real-time exploration of any dataset, and so on.
- ➢ **Sharing:** It provides safe and secure way to share the data. Also, allows connection to multiple data sources as well as there's simplicity in importing and exporting large amounts of data.

What do you see as the strengths of the software?

Following are the strengths of the Alteryx Designer and Tableau Public:

1. It saves a lot of time as the use of statistical language such as R/Python to implement K Means Clustering does not require the writing of code pages.
2. Error identification and rapid data flow and working flow are simply achieved.
3. Tableau Public simplifies data analysis that provides customers with a model of storytelling.
4. Helps visual analysis, discovery, and participation.
5. Tableau can easily manage millions of rows of data with an ability to present live data with links to other data sources.

What do you see as challenges to using the software?

Following are the challenges of the Alteryx Designer and Tableau Public:

1. The processes might be difficult to construct depending on the business environment, while using macros.
2. Not useable when numerous users attempt to use the same workflow as unstable and often crashes on a server.
3. Tableau mostly focuses on visualization and cannot cope with uncleaned data correctly.
4. A good data clearance is necessary in the underlying database for the optimal usage of Tableau.
5. The schedule does not allow the automatic feedback option.

After using the software, do you recommend that businesses spend their resources on this product?

Yes, we would highly recommend using Alteryx Designer and Tableau Public for businesses to spend their resources on.

If not, what alternative do you recommend should be explored instead?

No, we would not recommend any alternative to Alteryx Designer and Tableau Public.

Future Developments

What are some foreseeable future developments of your software?

Alteryx Designer has stepped into the intuitive, cloud native data prep, blending, and analytics solution that is accessible to anyone in the enterprise who interacts with data. There is a strong focus on Cloud Computing and Artificial Intelligence by Alteryx.

Tableau is heavily investing in AI, Advanced Analytics. It aims at performing smarter, guided decisions with cutting-edge, augmented analytics in a no-code required, rapid iteration environment.

What are some key trends to monitor in this software category?

Some of the key trends to monitor in the category of Alteryx Designer are:

1. Augmenting AI for intelligent data ingestion and processing.
2. Cloud data preparation has taken the center stage.

Some of the key trends in the category of Tableau Public are:

1. Incorporating AI and Machine Learning techniques to identify patterns using. powerful algorithms for massive datasets.
2. Shift in focus from visualizing the data to augmented and virtual reality, to make interaction with data more cognitive.
3. Increasing focus on Real Time Data Visualization.