**TASK1**

eBay needed a Big Data solution to accommodate and generate its explosive data growth as the world’s largest online marketplace which enables the buying and selling of practically anything. The solution was required so that eBay’s increasingly copious amounts of data can be processed at blistering speeds and rapidly analyzed to generate useful insights for its customers. Moreover, there was a need for a solution to be able to perform rapid analysis on a broad assortment of the structured and unstructured data it captured.

The challenge faced by eBay was supporting and managing data at an extreme scale. eBay was experiencing rapid explosive data growth and needed a solution that did not have typical bottlenecks (the process(es) with the lowest output which may be a hindrance to agile processing and analysis of huge amounts of data), scalability issues (which may arise when apps fail to scale up or down when handling large volumes of data) and transactional constraints associated with common relational database approaches. To accommodate eBay's explosive data growth, its data centers perform billions of reads and writes each day. Due to the increasing demand to process data at blistering speed, eBay needed a solution to perform rapid analysis on a broad assortment of the structured and unstructured data it captured.

The solution proposed to address these challenges was to Integrate real-time data and analytics, which incorporates a scale-out architecture that enables eBay to deploy multiple DataStax Enterprise clusters across several data centers using commodity software. NoSQL technologies, specifically Apache Cassandra and DataStax Enterprise and its integrated Hadoop analytics were sought out by eBay’s Big Data requirements. The solution which is built on Apache Hadoop has a redundant architecture with multiple clusters which ensure the replication of data across several nodes which enable linear scalability. In a scalable architecture, resource consumption should be increased linearly (or better) with load, measured by user traffic, data, etc, therefore, it is a prerequisite for functionalities within eBay’s operations of handling Big Data. Furthermore, the Big Data solution did not integrate into a single Big Data center infrastructure. These components work together to enable eBay to store, process, and analyze copious volumes of data efficiently at extremely high speeds.

As a result of incorporating this big data solution, eBay is now able to process massive amounts of data more cost-effectively at very high speeds, at very high velocities, and achieve far more than they were able to with the other higher cost propriety system they had been using. Currently, eBay is managing a sizable portion of its data center needs that is, 250TBs+ of storage within the Apache Cassandra and DataStax clusters. However, according to the case study, eBay is now able to quantify the social data it displays on its product pages through load balancing and application availability. The Cassandra distribution in DataStax Enterprise stores all the information needed to provide counts for “like”, “own,” and “want” data on eBay’s product pages. It also provides the same data for the eBay “Your Favourites” page that contains all the items a user likes, owns, or wants with Cassandra serving up the entire “Your Favourites” page. As expressed, eBay provides this data through Cassandra’s scalable counters feature. Load balancing helped eBay meet its big data requirements with extremely fast data handling and application availability aspect. Balancing incoming user load and eliminating application downtime enabled the buying and selling of anything in the online marketplace. eBay architects had the flexibility they needed to design a system that enables any user request to go to any data center with a DataStax Enterprise cluster spanning those centers. eBay is also able to perform high-speed analysis with the ability to maintain a separate data center running Hadoop nodes of the same DataStax Enterprise ring. Another result obtained from using the big data solution is the use of the Hunch “Taste graph” for eBay users and items to provide customers’ recommendations based on user interests. All events that take place on eBay’s website are captured and stored as a graph in Cassandra. Furthermore, this new architecture (DataStax Enterprise) serves a variety of time-series use cases in which processing high-volume, real-time data is the foremost priority. Finally, the DataStax Enterprise enabled eBay to process high-volume, real-time mobile notification logging and tracking (every time eBay sends a notification to a mobile phone or device, it is logged in Cassandra), fraud detection, SOA request and response payload logging, as well as RedLaser (another eBay sister company) server logs and analytics.

Using data centers means there is instant scalability and the facilitation of smooth transactions for customers. They have the infrastructure in place to scale up storage capacity quickly and easily, without compromising security or performance. With proper configuration, data centers can instantly handle the changes in web traffic and engagement to the needs of the company without interrupting customers. Data centers provide unprecedented levels of security thus reducing burdens for eBay. They are equipped with the latest high-security mechanisms and tools to securely handle and store data thus preventing cybercriminals from infiltrating and disrupting operations within eBay’s online applications. They have robust physical and digital security measures in place to protect customers’ data from theft, unauthorized access, and natural disasters (grouponeit.com, n.d.). Data centers give eBay a sense of reliability and increased uptime. They are equipped with a reliable offsite cloud as well as redundant backup and power supplies without the glitches of portable technology to ensure that eBay’s data is safe in the event of natural power outages and that services are always available to its customers. They also have strict environmental controls to protect the data from dust, water, and extreme temperatures.

A single point of failure is a flaw in the design, configuration, or implementation of a system, software, or any non-redundant component that may pose a potential risk that would eventually cause the entire system to fail (SearchDataCenter, n.d.). In this case, eBay uses an integration of multiple systems which are critical in running the platform hence, if any of those individual components malfunction, the eBay platform may stop working. There is no likelihood of a ‘single point of failure’ SPOF occurring due to the use of a redundant (scale-out) architecture with multiple DataStax Enterprise clusters which ensures replication of data across several nodes. The solution has linear scalability which is found in the DataStax Enterprise Hadoop’s distributed architecture which allows nodes to be added in the Cassandra Ring where necessary hence reducing any risks of a single point of failure occurring.

The Balanced Scorecard provides a framework for integrating qualitative measures into eBay’s operations. The business process perspective views eBay’s organizational performance through the lenses of efficiency and effectiveness related to its services and operations to answer the question “What must we excel at?”. Some nonfinancial key metrics which are assessed to identify how well the business is running include Downtime or Uptime, Operating efficiency, Equipment utilization, and Quality control. eBay should attempt to identify and manage its core competencies, and the critical technologies needed to ensure continued market leadership. After doing so, it should decide what processes and competencies they must excel at and specify measures for each (Kaplan and Norton, 1992). Continuous tracking and improvement initiatives such as lean six sigma can also help eBay enhance its operations to ensure ongoing and continuous excellence of Its website. This perspective helps eBay to identify areas in its operations and processes that must be improved to create value for stakeholders and customers and to keep them satisfied.

The financial perspective is about the measurement of eBay’s success in terms of financial performance. It also looks at the effective use of resources (including financial resources) within an organization. This perspective is the strongest lagging perspective, which means that it is the outcome/output of the other 3 perspectives in the balanced scorecard. Furthermore, the financial perspective looks at top-level financial objectives to answer the question, “How do we look to our shareholders? and “How does your organization generate financial returns?”. It takes the viewpoint of company shareholders and typically looks at Revenue, Expenses, ‘Return on Investment” ROI and profit margins, and other key performance indicators. By measuring and tracking these KPIs, eBay can identify areas that are not producing revenue as expected and make appropriate adjustments when re-strategizing such as creating corrective plans according to recommendations from reports.

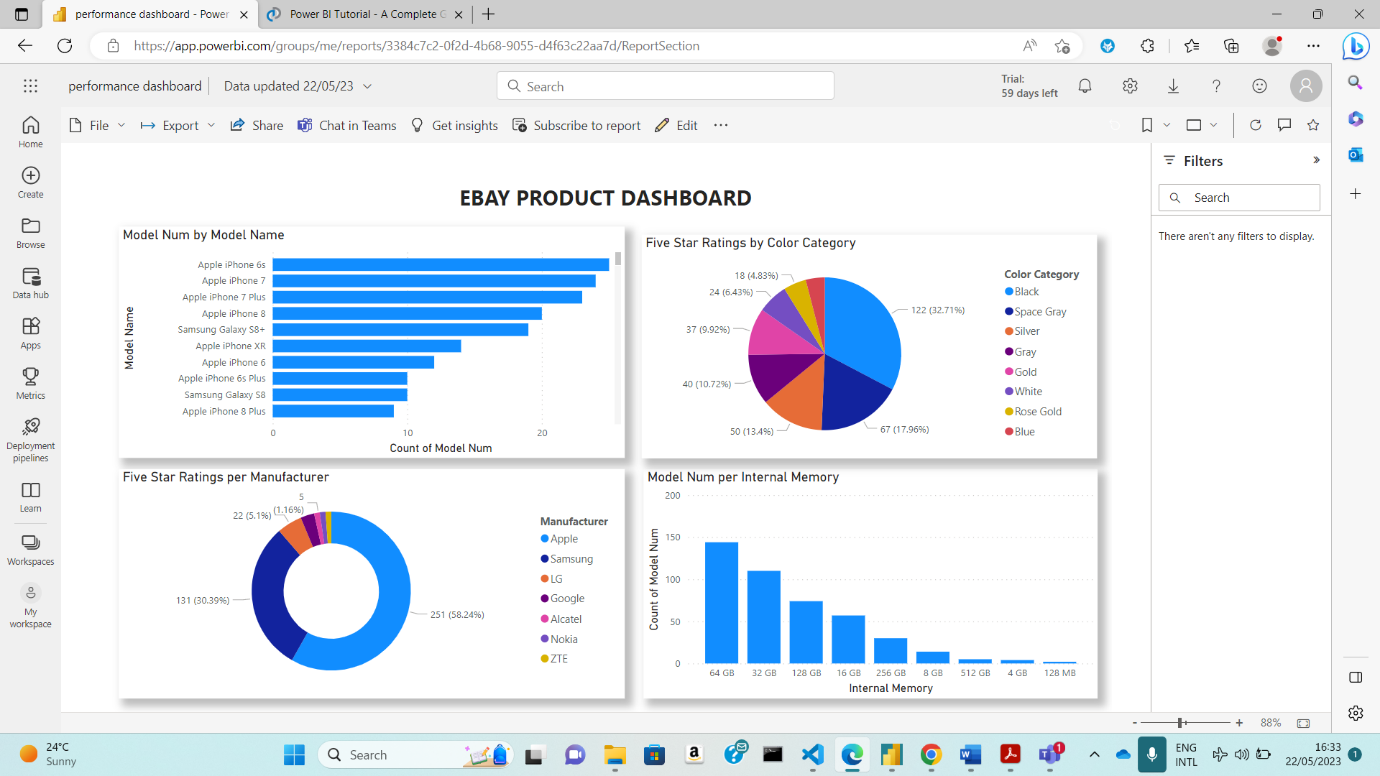
Predictive analytics is transforming data into future insights. It is a category of data analytics aimed at making predictions about future outcomes based on historical data and analytics techniques: **statistical modelling** and machine modelling. Predictive analytics has tools and models to use past and current data to reliably forecast trends and behaviours. To gain insights from all sorts of structured and unstructured data within eBay’s data repositories, statistical techniques such as linear regression models, neural networks, and decision tress can be used to define patterns and make predictions about future events. eBay can consider using **machine learning** **and data mining techniques** to analyze large datasets and make predictions about future behaviour, product demand and pricing trends. This can help optimize its pricing strategies, improve inventory management, identify new revenue opportunities. Some modelling techniques use predictive learnings to help eBay make additional predictive insights.

Text analysis can help with analyzing text written in reviews by customers on eBay’s website or documentation to understand its purpose (examine text and find trends that can enable eBay to take strategic actions). It can offer incredible insights into media sentiment, workforce performance, competitor intelligence and more. This is at least partly because social media and the public internet has created a huge body of text that can be automatically analysed and understood. **Sentiment analysis** is a step within text analysis which allows eBay to computationally identify and categorize user expressions in a piece of text format be it in customer reviews, comments, or conversations about eBay in general (Wonderflow.ai, 2019). social media analysis uses data collected from social media networks. The social media impact on eBay’s website can be measured through **descriptive analytics** ( which uses statistics to identify activity characteristics and trends), **social network analysis** ( which follows a link between followers, friends, and other organizations to identify connections of influence as well as the biggest sources of influence) and advanced analytics/ **Web Analytics** to examine online content that would not be revealed by casual surveillance and track and gather information about traffic levels and user behaviour on the website to improve user experience. **Competitive Analysis** is another strategy used in social media analysis that can help eBay understand what competitors are doing and how customers are responding to what they are doing. This can help eBay to follow along with trends within the online market.

**TASK2**

The uploaded dataset has raw data. This data has to be cleaned before building the report / dashboard. To check the quality of data on a dataset that has “missing Data is Yes”, open Power Query Editor in Power Bi desktop to transform the imported data (select transform data). In the Power Query Editor, the data in selected query displays in the middle of the screen and, on the left side, the Query pane lists the available queries/ tables. All steps taken to shape the data are recorded and each time the query connects to the data source, it automatically applies the steps to the dataset. Remove unnecessary columns and filter rows to remove inconsistencies. To remove unwanted columns, select the columns to be removed, then, on the Home tab, select Remove columns. If there are repeating rows and missing data, replace the function to replace all the missing values and standardize the data. Check for other inconsistencies and apply certain changes such as text filters.

Uploading dataset onto power Bi service: Download the dataset from Kaggle (source for dataset). Save this dataset to OneDrive/ Student Drive. Right click on download dataset>Extract files to have an Excel.csv file. To upload or retrieve the dataset, open the Microsoft BI application. From the Home ribbon, select Get Data> More. The Get Data window appears, showing the many categories to which Power BI Desktop can connect. Select Excel sheet from Student Drive, (select the dropdown arrow for file type and select “All files (\*,\*)” so that the excel sheet can appear options to be loaded). View the dataset and then press upload data.

Building the dashboard: After cleaning the dataset in Power Query Editor apply changes to the dataset and close the Power Query Editor. Back on the Power Bi home window, select report view, select visuals/ graphs, and apply filters or select the data to be shown on the graph/visual. Looking at the data that I wanted to focus on, I selected a pie chart to present data as a percentage, adding up to 100. The pie chart to represent the percentage of five-star ratings for the phone model color categories and a donut chart for show five-star ratings for the selected manufacturers. Bar and column charts were also part of the report to compare values against a particular category. The clustered bar chart compared model names against model number and the clustered column chart compared model number against internal memory. 

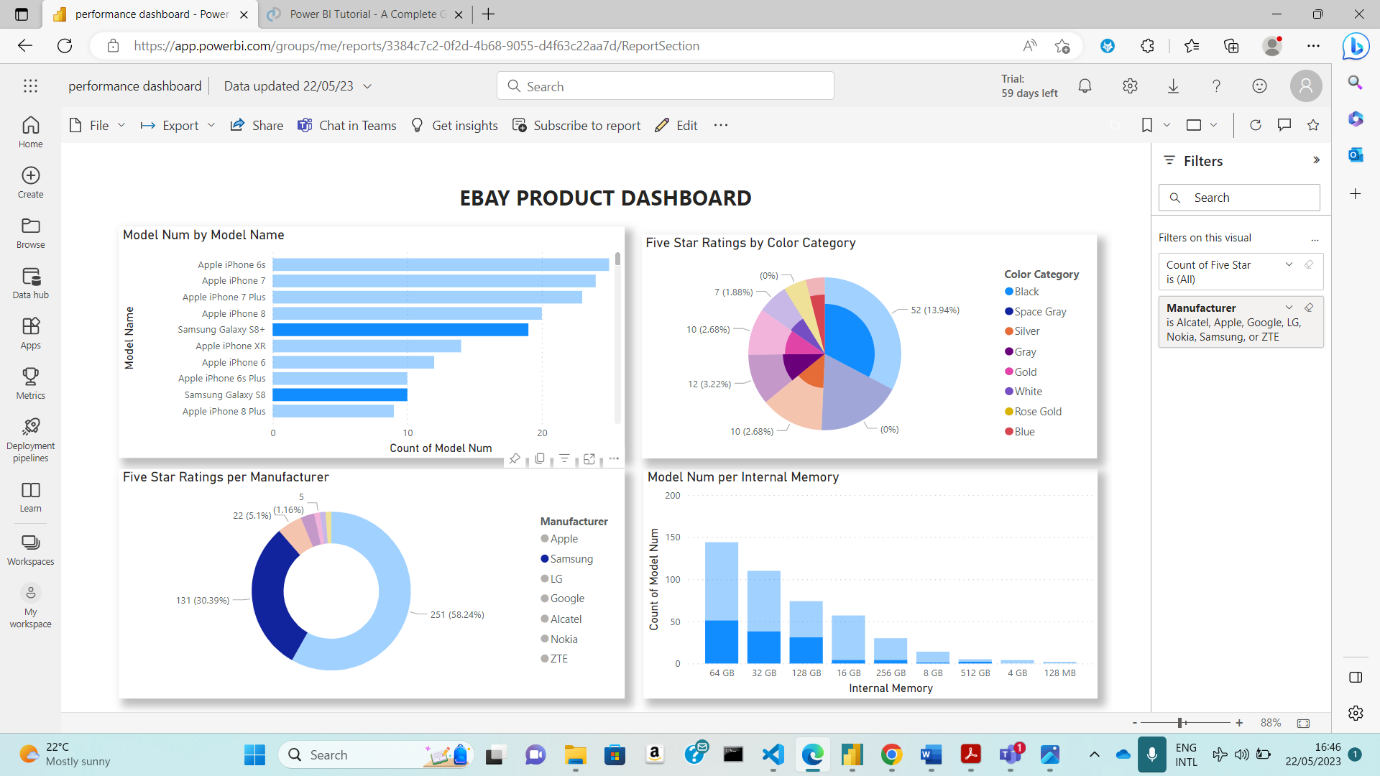
To share, the report, in the home ribbon, select publish report.

This will publish the report to the online power Bi app in “My Workspace”. Select share and it will give options to share the report. In the send link dialogue, the option to copying the sharing link or share it via Outlook, PowerPoint, and teams to people in my organization, people with existing access and specific people (lecturer or tutor). The most suitable would be to share the report with specific people by directly sending the link to the lecturer or tutor by entering their name or email address, optionally type a message, and select send.

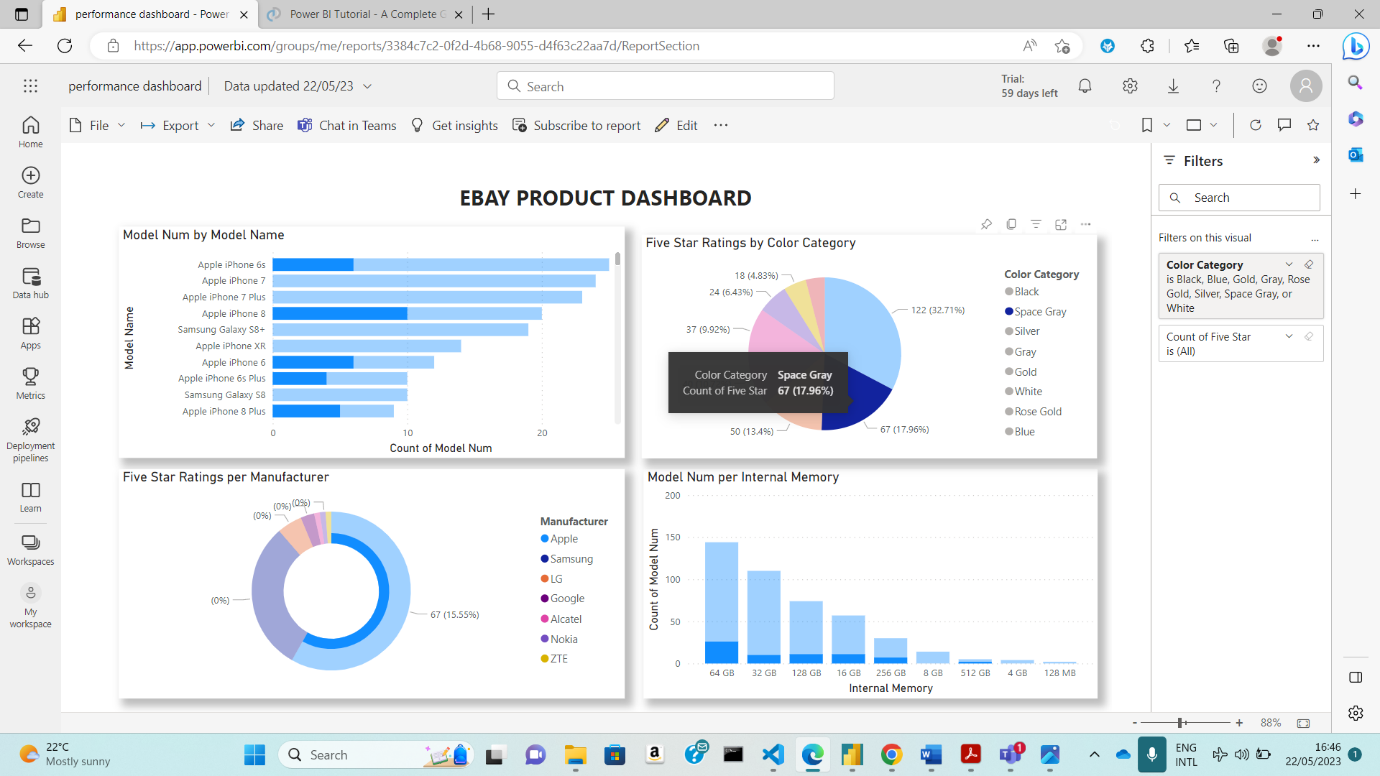
Samsung and Apple have noticeably more model numbers. Most people prefer to buy phones from Apple and Samsung since they have the highest percentage of five-star ratings as seen on the donut chart and these manufacturers have the highest number of models.

A screenshot of a computer

Description automatically generated



There may be a possibility that the quantity of models being produced affects the ratings or even a possibility of a small but significant limitation to buying from these brands.

The color of phones preferred is influenced by the manufacturer’s phone models. 

Nokia is the only manufacturer which produces phones that have an internal storage of 128MB.

A screenshot of a computer

Description automatically generated

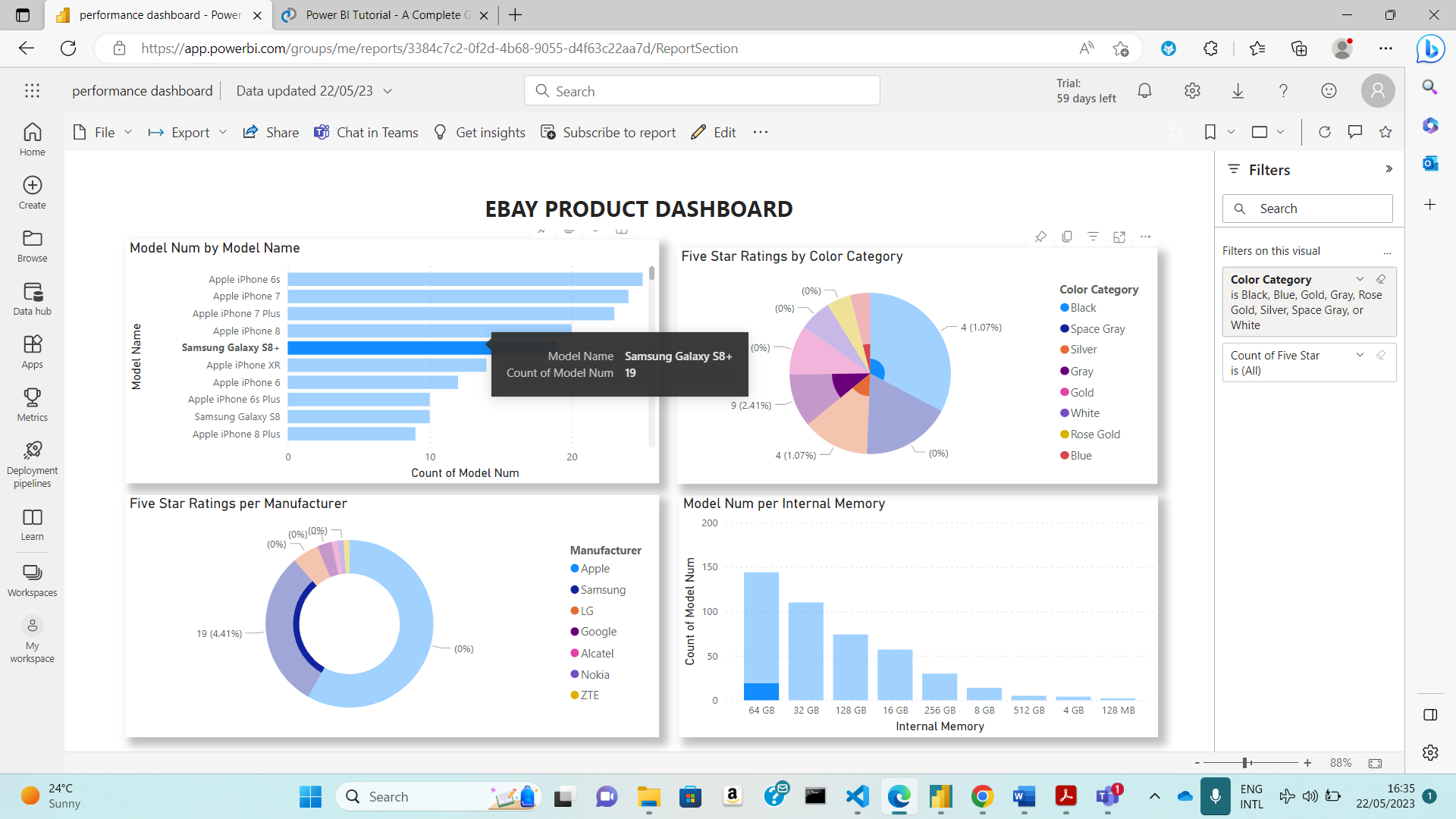
The five-star ratings for Nokia are also very low because very few people prefer a brand that produces phone models providing very little storage. Samsung has the highest number of models with an internal storage of 64GB.

A screenshot of a computer

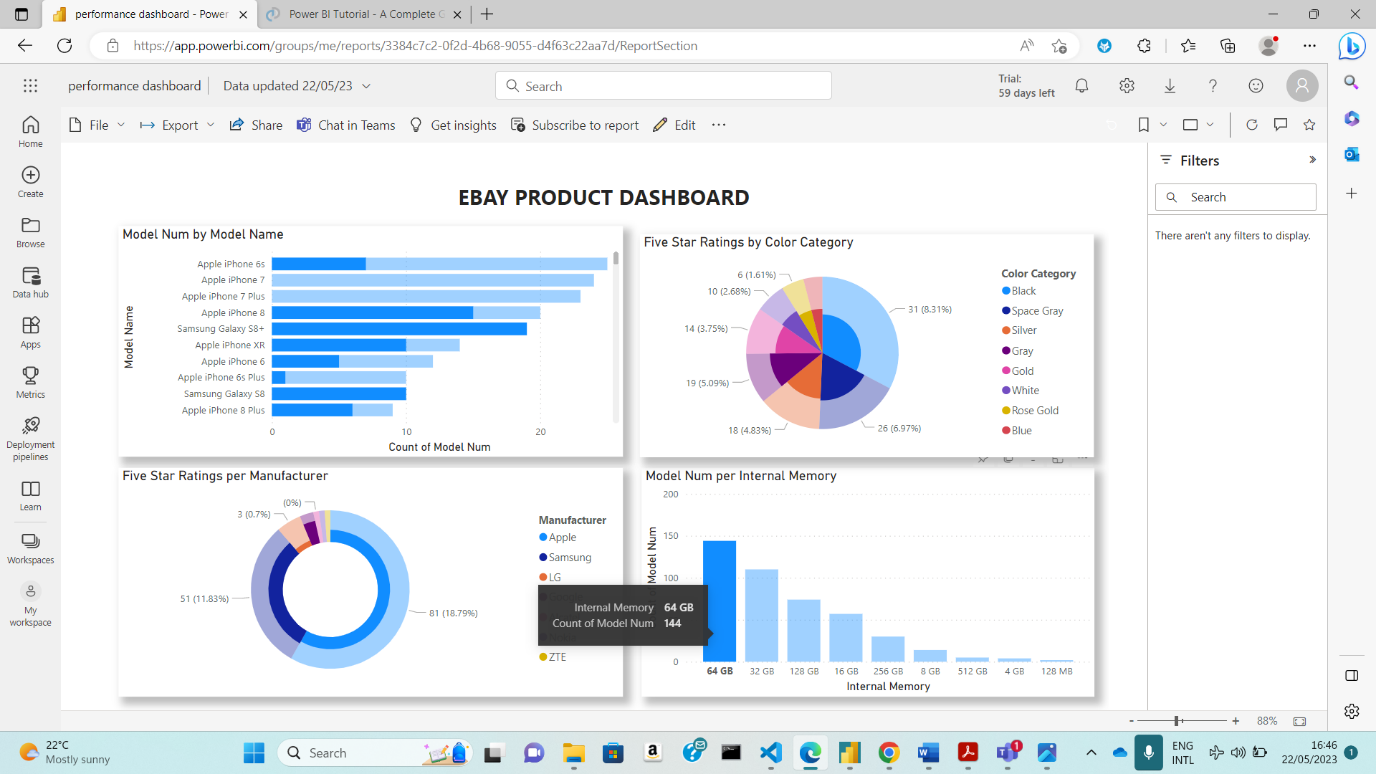
Description automatically generated

This also influences the ratings. The illustration can be oriented by using the axis and legends used in the graphs.

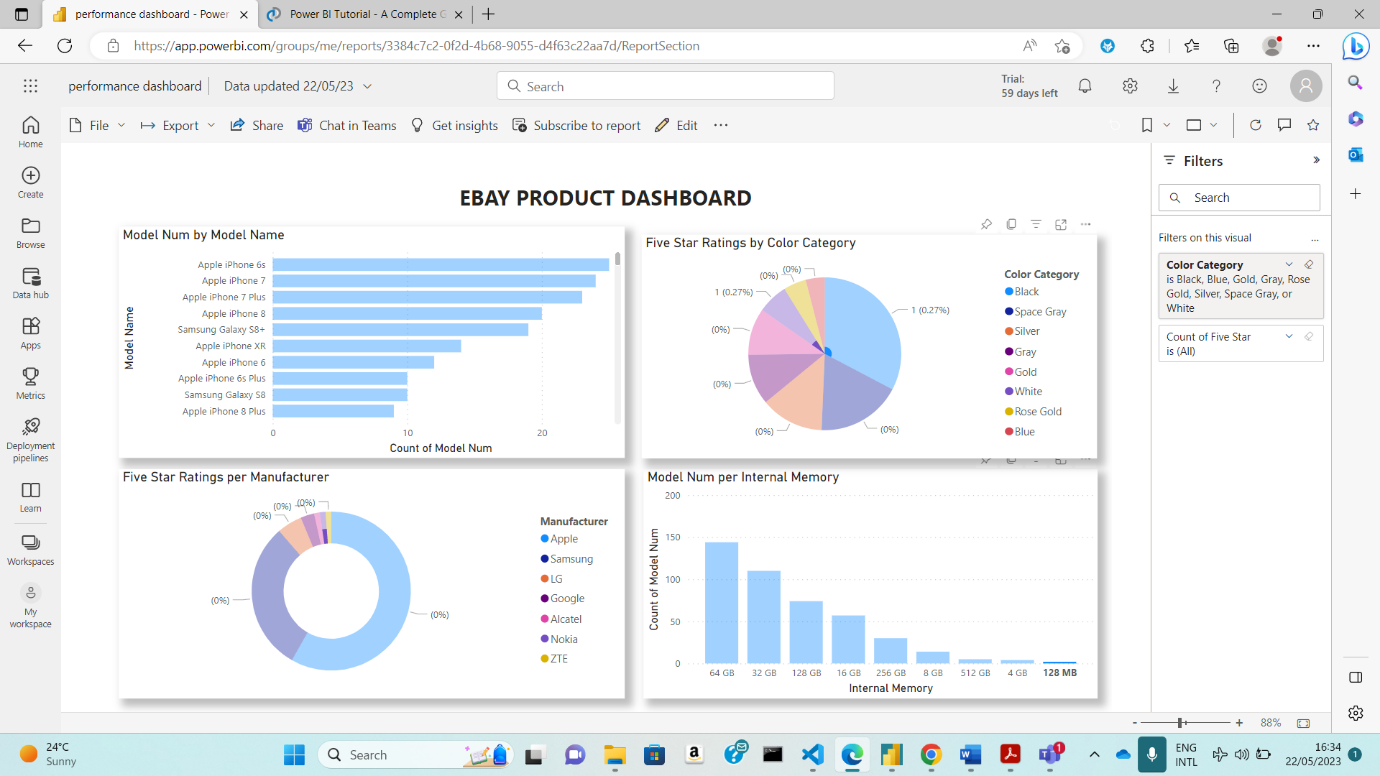
From the clustered bar chart which compared the number of models against their names, the results show that Samsung Galaxy S8+ had the highest number at 11, followed by Apple iPhone 7 and Samsung Galaxy S9+, which tied second at 7.



Across all 145 Model names, the count of model numbers ranged from 1 to 11. The clustered column chart looks at the number of models per each internal memory stated.

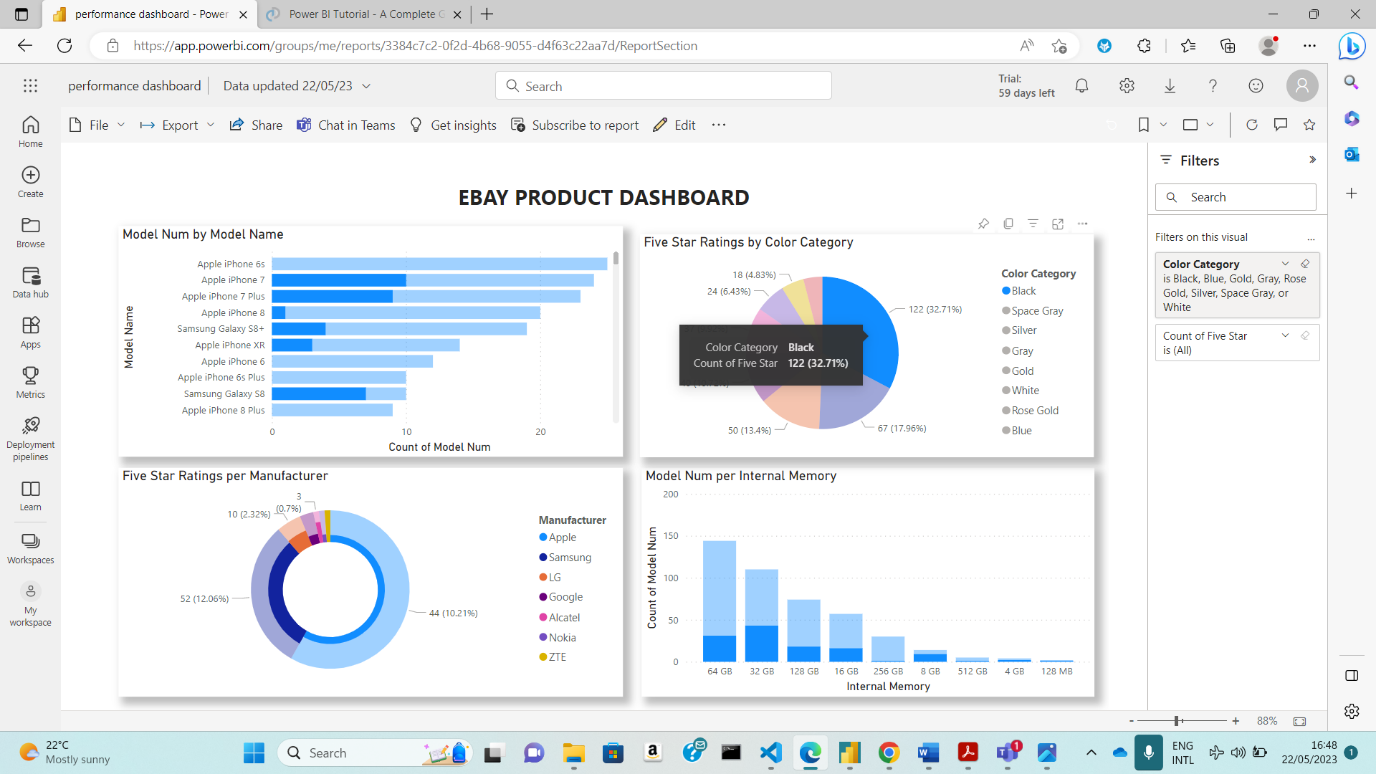


The highest number of models which is close to 150 had an internal memory of 64GB followed by 32GB with numbers just above 100



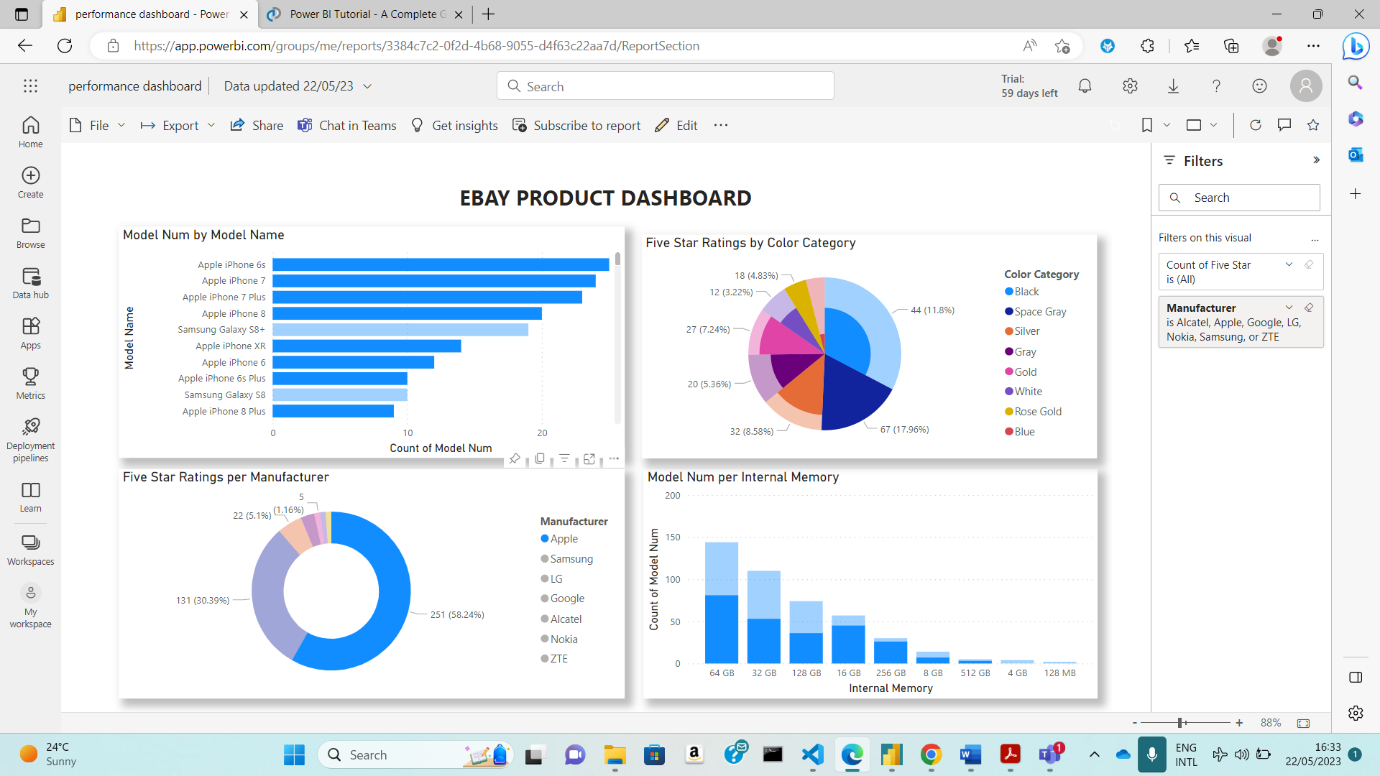
and the least number of models had an internal memory of 128MB.

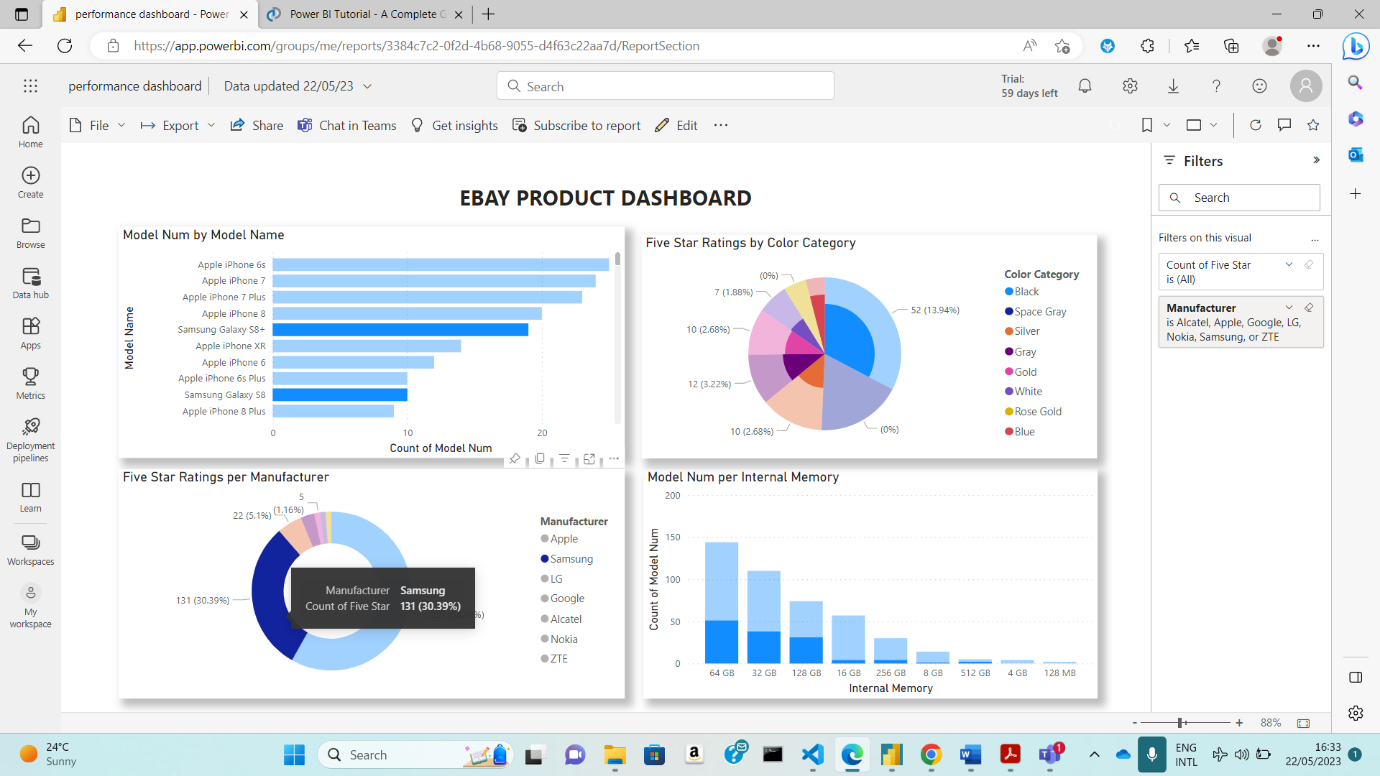
The pie chart shows that the most preferred color of phones is black, which has a high percentage of 32.1% followed by space Gray.



Gold, Gray, and silver fall within the same range of five-star ratings, the segments only differ with a percentage of 10% or below.

The donut chart shows that apple and Samsung have the most five-star ratings with Alcatel, Nokia, and ZTE with the least number of five-star ratings.





If one bar, segment, or component is selected in a graph, other graphs with change to show how they interact or relate with each other. For example, if I select the Apple segment in the donut chart, the other 3 graphs with highlight the parts where iPhone will fall within those graphs according to the respective axis. Therefore, it is possible to define and describe the graphs simultaneously.

In conclusion, all 4 graphs are linked and can be used to determine what type of models manufacturers produced according to what customers prefer. They can also help manufacturers to focus on producing models with specific internal storage based on what customers prefer instead of producing models which aren’t liked or bought so as to save resources and sell efficiently. Manufacturers can also use this information to focus on improving the quality of the most liked models. In other words, these graphs can be used in sentiment analysis (by using the ratings or customer reviews) and competitive analysis.

**REFERNCE LIST**

Kaplan, R.S. and Norton, D.P. (1992). *The Balanced Scorecard—Measures that Drive Performance*. [online] Harvard Business Review. Available at: <https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2>.

Sharda, R., et al (2018) Business Intelligence, Analytics and

Data Science, A managerial perspective, 4th Edition, Pearson education.

www.ibm.com. (2014). *What is a single point of failure?* [online] Available at: https://www.ibm.com/docs/pl/zos/2.1.0?topic=data-what-is-single-point-failure [Accessed 22 May 2023].

Lopez, M. (n.d.). *How eBay Uses Big Data and Machine Learning to Drive Business Value*. [online] Forbes. Available at: https://www.forbes.com/sites/maribellopez/2016/10/04/how-ebay-uses-big-data-and-machine-learning-to-drive-business-value/?sh=39afe46b1f35 [Accessed 22 May 2023].

grouponeit.com. (n.d.). *Benefits of a Data Center*. [online] Available at: <https://grouponeit.com/benefits-of-a-data-center>.

literature, G.W. is a content creator within the marketing team at B.S. loves creating actionable content in different formats to help others achieve more project success G. spent far too long at university studying E., Learning, W.I. a L.-L.L. of, time, upskilling I. her free, Gym, S.E. a C.S. at the, Book, T. into a G. and dog, walking the beautiful G. coastline with her (n.d.). *How to Create Reports with Power BI Service*. [online] BrightWork.com. Available at: https://www.brightwork.com/blog/power-bi-create-report [Accessed 22 May 2023].

SearchDataCenter. (n.d.). *What is a single point of failure (SPOF) and how to avoid them?* [online] Available at: https://www.techtarget.com/searchdatacenter/definition/Single-point-of-failure-SPOF.

‌mihart (2023). *Understand how visuals interact in a report - Power BI*. [online] learn.microsoft.com. Available at: https://learn.microsoft.com/en-us/power-bi/consumer/end-user-interactions [Accessed 22 May 2023].

Wonderflow.ai. (2019). *Everything you need to know about text analysis - Wonderflow | Wond...* [online] Available at: https://www.wonderflow.ai/blog/text-analysis-examples#:~:text=Text%20analysis%20is%20really%20the.

‌

‌www.proquest.com. (n.d.). *How eBay Uses Data and Analytics to Get Closer to - ProQuest*. [online] Available at: https://www.proquest.com/docview/1438813225?pq-origsite=gscholar&fromopenview=true [Accessed 22 May 2023].

Team, D. (2018). *Power BI Tutorial - A Complete Guide on Introduction to Power BI*. [online] DataFlair. Available at: https://data-flair.training/blogs/power-bi-tutorial/.

‌ https://www.kaggle.com/datasets/promptcloud/ebay-product-listingdataset?resource=download&select=marketing\_sample\_for\_ebay\_comebay\_com\_product\_details\_\_20200901\_20201031\_\_30k\_data.csv

‌

‌

‌

‌