

MALEK AYMAN MOHAMAD DAHY (A20EC3020)

ABDULLAH YASER ABDO ALI ALNADHARI (A21EC0243)

EIKIESHA DEVI A/P SUBRAMANIAM (A21EC5015)

MOATAZ ABDO MOHAMMED AL-SHATER (A21EC4011)

MUHAMMAD WAFFI QAYYUM BIN DIN (A21EC0097)

Sales Statistics for global E-Commerce business

Trends of data analytics on different sectors and reflection of these trends

Data analysis is a set of procedures aimed at extracting relevant and useful information from large and diverse sets of data acquired from many sources and of variable sizes. Browsing activity and online purchases are two examples of data analysis.

The role of big data in data analytics, which involves the use of the 10 V's, which had indeed been mentioned. Volume, speed, diversity, authenticity, value, validity, variability, location, vocabulary, and vagueness are all factors to consider. Descriptive analytics, diagnostic analytics, productive analytics, and preceptive analytics are the four main forms of data analytics that have been mentioned as well.

The process of Data Analysis consists of 4 steps. The first one is splitting the data by multiple criteria which can be categorical or numerical such as age, gender, etc. The second step is collecting the data, which can be collected from different resources. The third step is organizing the data, the data must be organized in order to be examined. The fourth and final step is cleaning the data, this makes sure that there is no overlaps or mistakes made.

A Data Analyst has five roles. The first is to study information, then clean it from noise, then assess the quality of data and its sources, build scenarios for machine learning and finally supervise the procedures.

Data scientist works on multiple data operations, while the data analyst works on data cleaning and transforming. Data scientist deals with several fundamental data procedures, while data analyst is limited to small data and static inferences. Data scientist manages structured and unstructured data, while data analyst handles structured data only. Data scientist has mathematical, algorithmic and statistical knowledge, while data analyst possesses solving and basic statistical skills. Finally, Data analyst is an experienced in SAS,

Python, Hadoop, TensorFlow and Spark, while Data analyst is skilled in Excel, SQL and Tableau.

MICROSOFT POWER BI

Microsoft Power BI is a cloud-based, business-intelligence application from Microsoft that offers analytics services such as data analysis and visualisation. Power BI provides a platform for users to be creative with reports and analytics as there are numerous capabilities that users can readily access to visualise and manipulate data. Users are able to update their report data without having to create a new report every single time. Many companies prefer Microsoft Power BI because of its low cost compared to other software while allowing users to manipulate data in a variety of ways. For example, data can be transformed by direct modification(front end) utilising the application's provided features or by accessing the back end using (Mquery, Dax). As a result, Power BI is crucial to modify the data using a variety of approaches immediately and also visualizing data easily.

Steps to use Power BI :

Step 1 : Connect to users 'data by uploading or importing any type of files(excel, csv, sql etc.) into Power BI.

Step 2 : Data preformatting where users can edit the files to meet their specific requirements by renaming columns, eliminating rows and so on.

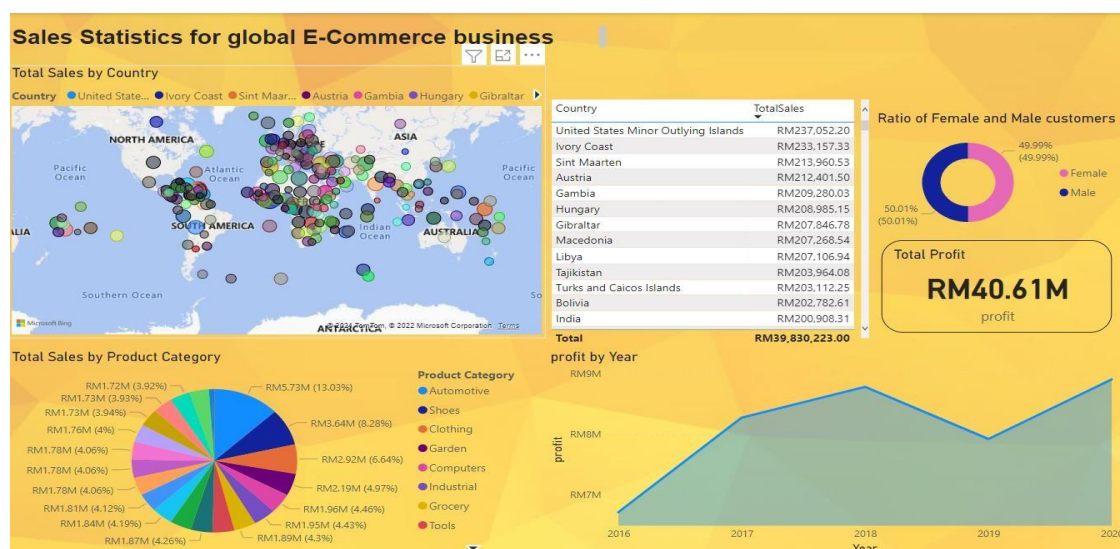
Step 3 : Data Modelling where users will create or enhance data by adding relationships, calculations and other features. Query can also be written to improve data visualisation and analytics later on.

Step 4 : Data visualization which users will experiment the data with different visualisation types. Reports are able to be created using visual tools and without writing any code.

Step 5 : Using Power BI's supplied services, data reports can be published. Users can safely export and publish their reports (via QR code, Power Point, embed in public web, etc.) as well as set up automatic data refresh, which gives the team with near-real-time trend and indicator analytics.

In conclusion, Microsoft Power BI is a powerful software that offers a variety of services to assist users in visualising, manipulating, and creating an excellent report. The users can create a report without having to write a single line of code. This will save a lot of time without compromising the quality of the result. Power BI is also significantly less expensive than many other software that provide similar services, such as Tableau. Therefore, many businesses prefer to use Power BI over other software

Dataset Narrative: It was decided that we will talk about e-commerce and the percentages of the profits it sellers got from it especially with the technology rapidly improving and with the recent pandemic in which people were only able to purchase goods online. In the beginning, we found a huge dataset talking about the world's e-commerce from the period 2016-2020. We firstly started by removing the unnecessary information such as ordering hours, and dates. In addition, we also checked for any duplicates with the exact same entities. We then cleaned data from spelling errors in countries and organized data from highest to lowest in profits and countries and which gender ordered more. We used some formulas to calculate the information we wanted to get like profits per year. After all the organizations to the excel file, we imported it to Power BI in order to make the proper relations and visualizations.



Discussion: The stats showed that throughout the past five years e-commerce is steadily increasing with the technology simultaneously improving, however, in 2020, the numbers significantly rocketed because the whole world was affected by the pandemic and almost all purchases were done online. With regards to the products sold, an unsurprising fact appeared as Automotive products were the highest to sell with the majority of almost 5 million ringgits. It was then closely followed by the fashion products with shoes selling a bit higher than clothes. Top 4 and 5 were reserved for garden and computer products, respectively. Apart from products, the countries using e-commerce showed 2 significant details as the United States was the highest country to purchase which should not result in any surprises, however, the surprising fact was the Ivory Coast came second on the list. As with genders, the stat showed that both men and women have close numbers of purchases with men being merely better than women. To sum up, this shows that e-commerce is just about to become more and more vital in the way individuals purchase goods.

Reference: Gbenga, O. (2022, January 12). E-commerce dataset. Kaggle. Retrieved January 13, 2022, from <https://www.kaggle.com/onigbenga/ecommerce-dataset>