ASSIGNMENT 2

**EXCEL**

**HISTORY**

Excel is a widely used spreadsheet program that offers basic data analysis capabilities that allows users to organize and manipulate data, perform calculations, create charts and graphs, and conduct basic statistical analysis. Excel can act as a simple database program to store, search and retrieve data and also offers a range of uses, features and functions for data analysis.

**Founders**

Excel was developed by Microsoft.

**Advantages**

The following Excel uses can be utilized by professionals from various career fields;

Excel provides a simple and efficient way to enter, organize and store data in a tabular format. It allows users to perform various data analysis tasks like sorting and filtering data, calculating statistics, creating charts and graphs to visualize data patterns. It offers a variety of chart types and formatting options to create visual representations of data for presentations and reports. It enables users to create dynamic reports and interactive dashboards by combining data, charts and visuals on a single sheet. Excel is commonly used for financial tasks like budgeting, forecasting, tracking expenses and creating financial models. In project management it can be used for project planning, scheduling, tracking progress and generating reports.

The following are the Excel features available for use;

Excel provides a vast library of built-in functions and formulas for performing calculations, such as Sum, Average, Count, IF, Vlookup and more. Conditional formatting allows users to highlight cells based on specific conditions, making it easier to identify trends, outliers or important data points. PivotTables and PivotCharts allow users to summarize, analyze and visualize large datasets quickly by dynamically rearranging and summarizing data. Data validation allows users to set rules and constraints for data entry, ensuring data accuracy and consistency. Data Import and Export enables users to import data from external sources like databases, text files and web pages, as well as export data to various formats. Collaboration and Sharing allows multiple users to work on a spreadsheet simultaneously and provides sharing options to collaborate with others. Excel’s What-If analysis tools, such as Goal Seek and Data Tables, allow users to explore different scenarios and understand how changes in variables impact the results.

Excel has a range of functions available for use which include;

Sum which is used to add up a range of numbers. Average which calculates the average of a range of numbers. Count which counts the number of cells that contain numeric values. IF which performs a logical test and returns different values based on the result. Vlookup which looks up a value in a specific range and returns a corresponding value from a different column. Concatenate which joins multiple text strings into one. Today which returns the current date. PMT which calculates the payment for a loan or investment with fixed payments and interest rate. MAX returns the maximum value from a range of numbers, while MIN returns the minimum value from a range of numbers.

**Demerits of Using Excel**

The disadvantages of using Excel include the fact that it offers basic data analysis capabilities. There are limitations on its versatility in comparison to other tools. It does not allow data sharing or collaboration among teams and there is lack of control and security. Excel is prone to human error and is hard to consolidate. Excel is unsuitable for agile business practices, is difficult to troubleshoot or test and it can't help us make quick decisions. Excel is not designed for collaborative work. Excel is vulnerable to fraud/corruption and difficult to manage advanced pricing rules.

**POWER BI**

**HISTORY**

Power BI is a business intelligence tool created by Microsoft that enables users to transform raw data into interactive visualizations and reports. It offers a user-friendly interface and allows users to perform data exploration, dashboard creation and data sharing. It is one of the four components of Power Platform.

**Founders**

Power BI was developed by Microsoft.

Advantages

Power BI has the following uses;

Data visualization that enables users to create interactive and visually appealing dashboards, reports and charts to present and explore data. Data exploration and analysis allows users to perform exploratory data analysis by filtering, sorting and drilling down into the data to uncover patterns, trends and outliers. Data modeling and transformation allows users to shape and transform data from different sources for analysis. Data collaboration and sharing enables users to share and collaborate on reports and dashboards with colleagues, stakeholders and clients. Real-time data monitoring allows users to connect to live data sources, enabling real-time monitoring and analysis of data as it updates.

Power BI has the following features;

Data connectivity allows users to connect to a wide range of data sources, including databases, spreadsheets, cloud services and APIs, to bring in data from multiple sources for analysis. Data preparation and transformation using Power Query, a data preparation tool within Power BI, enables users to clean, transform and shape data before loading it into the data model. Interactive visualizations such as bar charts, line charts, maps, tables and custom visuals, can be customized and interacted with to explore data in different ways. Advanced analytics like forecasting, clustering and quick insights allow users to uncover patterns and trends in the data. Data Analysis Expressions (DAX) is a language similar to Excel which users can utilize to create calculated columns and measures for advanced calculations and analysis. Power query editor allows user to transform and clean data using an intuitive interface, including features like data merging, filtering, splitting columns and applying transformations. Natural language query allows users to ask questions in plain language and get answers in the form of visualizations or data tables. Power BI Mobile provides users with mobile apps for iOS and Android devices, in order to access and interact with dashboards and reports on the go.

Power BI has the following functions;

Measures which enables users to create measures using DAX functions to perform calculations such as sums, averages, counts and more. Filters using various filtering functions users can narrow down data based on specific criteria or conditions. Aggregation functions like sum, average, max, min and count can be utilized to summarize data in visualizations. Time Intelligence functions like SAMEPERIODLASTYEAR, TOTALYTD and DATEADD, allow users to analyze data over different time periods. Statistical functions including correlation, standard deviation, rank and percentile allow users to perform advanced data analysis.

**Demerits of using Power BI**

Cost: Power BI is a premium service, and the cost might vary depending on the version and features needed. Steep Learning Curve: Particularly for individuals who are unfamiliar with Microsoft products or data analysis techniques, Power BI has a steep learning curve. Limited Data Sources: While Power BI offers connectivity to a variety of data sources, some specialized data sources or data types might not be supported. Online-only: Because Power BI is a cloud-based application, users need to be connected to the internet in order to utilize it. Performance: Power BI may take longer to process data or produce reports than other data analysis tools, depending on the size of the data set and the complexity of the study. Data Volume: Power BI has restrictions on the volume of data that may be fed into the system, despite the fact that it can handle massive data sets. The kind of Power BI licence being utilised determines how much data may be loaded. Data Complexity: Very complex data models are not intended to be handled by Power BI. Data Processing: When working with big data sets, Power BI’s lack of in-memory data processing can affect speed. Limited Visual Customization: Power BI offers a variety of visuals, but there are limits to how much customization is possible. Custom Code: Power BI’s functionality can be expanded using custom code, however writing and debugging this code can be difficult. Data Security: Although Power BI has strong security capabilities, the amount of data that can be secured is constrained. Compatibility Issues: There’s a chance that not all data sources and other applications will work well with Power BI.

**TABLEAU**

Tableau is a powerful data visualization tool that allows users to create interactive dashboards, reports and charts. It also enables users to connect to files, relational and Big Data sources to acquire and process data. The software enables data blending and real-time collaboration, which makes it unique.

**Founders**

Tableau was founded by Pat Hanrahan, Christian Chabot and Chris Stolte and its headquarters are in Seattle, Washington.

**Advantages**

Tableau has a range of uses that include;

Data visualization which enables users to create interactive and visually compelling charts, graphs, maps and other visualizations to explore and present data effectively. Data exploration and analysis allows users to easily explore and analyze data from various sources, apply filters, drill down into details and identify trends, patterns and outliers. Dashboards and reports enable users to create interactive dashboards and reports by combining multiple visualizations and elements onto a single canvas, providing a comprehensive view of data. Data storytelling enables users to create compelling data stories by combining visualizations, annotations and narratives to communicate insights and findings effectively. Collaboration and sharing allows users to share dashboards and reports with colleagues, stakeholders and clients, enabling real-time data collaboration.

Tableau users can utilize the following features;

Data connectivity to connect to a wide range of data sources, including databases, spreadsheets, cloud services and APIs, to bring in data from multiple sources for analysis. Drag-and-Drop interface allows users to easily create visualizations by dragging fields onto shelves and choosing visualization types. Interactive visualizations like bar charts, line charts, scatter plots, maps, treemaps, and more, allow users to customize visualizations, apply filters and drill down to explore data interactively. Data Blending and Joins allow users to blend and join data from multiple sources, allowing for integrated analysis and visualization of data from different databases or files. Calculated fields can be created by users using formulas, aggregations and logical expressions to perform custom calculations and analysis. Parameters allow users to create interactive elements that enable dynamic control over visualizations, such as adjusting filters or switching between different measures. Mapping and Geographic analysis allow users to plot data on maps, perform geospatial analysis and visualize location-based data effectively. Advanced analytics integration with statistical and analytical tools like R and Python, enables advanced analytics and predictive modeling within Tableau visualizations. Story points enable users to combine multiple visualizations into a coherent narrative or presentation, guiding viewers through a series of insights or findings. Tableau public is a free version that allows users to share visualizations and dashboards publicly on the web, fostering data sharing and collaboration.

**Demerits of Tableau**

 Poor Versioning is the main disadvantage of using Tableau as only recent versions support revision history and for the older one's package rolling back is not possible. No Automatic Refreshing of Reports as users don’t get an automatic option to refresh reports with the help of scheduling. Therefore, some manual effort required to update the data in back-end. Need Manual Effort as the parameters are inactive and only a single value can be selected using a parameter. You need to update it manually whenever the data gets change. Not a Comprehensive Solution as even if the Tableau Software is easy to use for BI application, still it doesn’t provide any platform for developing analytic applications that can be widely shared. No Version Control as once the dashboards and reports are published on the server you can’t get back to the previous levels of data in Tableau. It is not possible to go back and recover old data. SQL Knowledge is required in order to create rich and complex datasets from multiple data sources. A tableau is an excellent option for business users to play with the data once the dataset is created.

**PYTHON**

Python is a popular programming language known for its simplicity, versatility and extensive libraries and frameworks namely NumPy, Pandas and Matplotlib. These libraries provide extensive capabilities for data manipulation, statistical analysis, visualization and machine learning. It includes tools for performing various statistical tests, as well as linear regression and time series analysis. It makes it easy to conduct data cleaning and management, distinctly styled graphs, descriptive analysis and advanced analysis.

**Founders**

Python was founded by Guido van Rossum in the Netherlands.

**Advantages**

Python has the following uses;

Web development using frameworks like Django and Flask, to build scalable and feature-rich web applications. Data analysis and visualization using libraries like NumPy, Pandas and Matplotlib, which provide tools for handling large datasets, performing statistical analysis, and creating interactive visualizations. Machine learning and AI using libraries like TensorFlow, Keras and Scikit-learn which provide powerful tools for building and training machine learning models. Scripting and automation due to its simplicity and readability, it can be used to automate repetitive tasks, process files and interact with operating systems. Scientific computing using libraries like SciPy, SymPy and OpenCV which provide functionalities for scientific computations, symbolic mathematics and computer vision. Game development using frameworks like Pygame, which simplify game development by providing tools for graphics rendering, sound and user input.

Python has the following features;

Easy-to-Read syntax due to its clean and straightforward syntax, which makes it easier to write and understand. Large standard library that provides ready-to-use modules for various tasks such as file I/O, networking, regular expressions, and more. Cross-Platform compatibility, due to its availability on multiple platforms including Windows, macOS and Linux, allowing developers to write code that can run on different operating systems. Dynamically Typed which means that variables are not bound to a specific data type, allowing for easier development and prototyping. Object-Oriented Programming (OOP) support which allows developers to organize code into classes and objects for better code structure and reusability.

Python has the following functions;

Print(): displays output on the console. Len(): returns the length of a string, list or other iterable. Range(): generates a sequence of numbers. Input(): takes user input from the console. Type(): returns the data type of a variable or value. Srt(): converts a value to a string. Int(): converts a value to an integer. Float(): converts a value to a floating-point number. Max(): returns the maximum value from a sequence. Min returns the minimum value from a sequence.

**Demerits of using Python**

Speed Limitations due to the fact that Python code is executed line by line, and since Python is interpreted, it often results in slow execution. This, however, isn’t a problem unless speed is a focal point for the project. Weak in Mobile Computing and Browsers as while it serves as an excellent server-side language, Python is much rarely seen on the client-side. Besides that, it is rarely ever used to implement smartphone-based applications. The reason it is not so famous despite the existence of Brython is that it isn’t that secure. Design Restrictions since Python is dynamically-typed. This means that you don’t need to declare the type of variable while writing the code. While this is easy on the programmers during coding, it can raise run-time errors. Underdeveloped Database Access Layers when compared to more widely used technologies like JDBC (Java DataBase Connectivity) and ODBC (Open DataBase Connectivity), Python’s database access layers are a bit underdeveloped. Consequently, it is less often applied in huge enterprises. Simple - Python’s simplicity can indeed be a problem whereby users become too comfortable to learn another programming language.

**SQL (STRUCTURED QUERY LANGUAGE)**

SQL is a language used to manage and analyze relational databases. It allows users to query databases, extract and manipulate data, and perform calculations. Popular SQL-based software includes MySQL, Microsoft SQL Server and Oracle. It is a user-friendly language because it is easy to learn and understand.

**FOUNDERS**

SQL was developed in the 1970s by IBM researchers Raymond Boyce and Donald Chamberlin.

**Advantages**

The following are the uses of SQL;

Database Management as users primarily use SQL to manage databases, including creating, modifying, and deleting databases, tables and indexes. Data Manipulation allows users to retrieve, insert, update and delete data from databases using commands like SELECT, INSERT, UPDATE and DELETE. Data Definition which provides commands for defining the structure and characteristics of database objects, such as tables, views, indexes and constraints. Data Querying and Reporting which allows users to write complex queries to retrieve specific data from databases, perform aggregations and generate reports. Data Analysis which enables users to perform tasks like, filtering, sorting, joining and aggregating data to gain insights and extract meaningful information from databases.

The following are the SQL features;

Relational Database Management System (RDBMS) Support: SQL is designed for relational databases and is supported by most major RDBMS systems, such as MySQL, Oracle, Microsoft SQL Server, and PostgreSQL. Declarative Language: SQL is a declarative language, meaning users specify what data they want, and the database management system determines how to retrieve or manipulate the data. Schema-Based Structure: SQL databases use a predefined structure called a schema that defines the organization and relationships between tables, providing consistency and data integrity. Data Integrity Constraints: SQL allows users to enforce data integrity through constraints like primary keys, foreign keys, unique keys, and check constraints, ensuring data accuracy and consistency. Transaction Management: SQL supports transaction management, allowing users to group multiple database operations into atomic units, ensuring data consistency and recoverability.

**SQL Functions**

SELECT: Retrieves data from one or more tables based on specified conditions. INSERT: Inserts new data into a table. UPDATE: Modifies existing data in a table based on specified conditions. DELETE: Removes data from a table based on specified conditions. DISTINCT: Retrieves unique values from a column. COUNT: Returns the number of rows or values in a column. SUM, AVG, MAX, MIN: Perform calculations on numeric data, such as summing, averaging, finding the maximum or minimum value.

GROUP BY: Groups rows based on specified columns for aggregate calculations. JOIN: Combines data from multiple tables based on specified conditions. WHERE: Filters data based on specified conditions.

**Demerits of using SQL**

Along with some benefits, the Structured query language also has some certain disadvantages:

Difficult Interface - SQL has a complex interface that makes it difficult for some users to access it. Partial Control - The programmers who use SQL doesn’t have a full control over the database because of the hidden business rules. Implementation - Some of the databases go to the proprietary extensions to standard SQL for ensuring the vendor lock-in. Cost - The operating cost of some SQL versions makes it difficult for some programmers to access it.

**R**

R is a programming language designed for statistical analysis and data visualization. It provides a wide range of packages and libraries for data manipulation, exploratory data analysis, statistical modeling, and graphical representation.

**FOUNDERS**

R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team. It was named R, based on the first letter of first name of the two R authors (Robert Gentleman and Ross Ihaka), and partly a play on the name of the Bell Labs Language S.

**Advantages**

**Uses of R**

Statistical Analysis: R is widely used for statistical analysis, hypothesis testing, regression models, time series analysis, and other advanced statistical techniques. It provides a comprehensive set of statistical functions and packages. Data Manipulation and Transformation which offers users powerful tools for data manipulation, including subsetting, merging, reshaping and transforming datasets. It allows for efficient data cleaning, handling missing values, and creating derived variables. Data Visualization which provides extensive capabilities for creating visualizations, including scatter plots, bar charts, line plots, histograms, and more. It offers flexible options for customizing and refining plots, making it suitable for exploratory data analysis and presentation-quality graphics. Machine Learning allows users to access a rich ecosystem of packages for machine learning and predictive modeling. It supports algorithms for classification, regression, clustering, dimensionality reduction, and more. It allows users to build and evaluate complex machine learning models. Reproducible Research: R is well-suited for reproducible research and data analysis workflows. It provides tools for creating reproducible reports, combining code and documentation, and facilitating collaboration.

**Features of R**

Extensive Package System which has a vast collection of user-contributed packages, called the Comprehensive R Archive Network (CRAN). These packages provide additional functions and capabilities for specific domains, making R highly extensible. Interactive Environment where users can execute commands and see immediate results. It supports an interactive command-line interface (CLI) and integrated development environments (IDEs) like RStudio. Data Structures offers users various data structures, including vectors, matrices, data frames, lists, and factors. These data structures provide flexibility and efficiency for storing and manipulating different types of data. Functional Programming paradigms allow users to write code in a concise and expressive manner. It provides functions as first-class objects and supports higher-order functions and functional composition. R has an effective data handling and storage facility as well as providing a large, coherent and integrated collection of tools for data analysis. R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

**R Functions**

read.csv(): Reads data from a CSV file into a data frame.

summary(): Provides summary statistics of a dataset, such as mean, median, minimum, maximum, and quartiles.

lm(): Fits a linear regression model to the data.

ggplot(): Creates visualizations using the ggplot2 package.

subset(): Filters a data frame based on specified conditions.

merge(): Combines two or more data frames based on common variables.

t.test(): Performs t-tests for hypothesis testing.

na.omit(): Removes missing values from a dataset.

apply(): Applies a function to rows or columns of a matrix or data frame.

predict(): Generates predictions from a fitted model.

**Demerits of using R**

It’s a complicated language. R has a steep learning curve. It’s a language best suited for people who have previous programming experience. It’s not as secure. R doesn’t have basic security measures. Consequently, it’s not a good choice for making web-safe applications. Also, R can’t be embedded in web browsers. It’s slow. R is slower than other programming languages like [Python](https://www.simplilearn.com/learn-the-basics-of-python-article) or MATLAB. It takes up a lot of memory. Memory management isn’t one of R’s strong points. R’s data must be stored in physical memory. However, the increasing use of cloud-based memory may eventually make this drawback moot. It doesn’t have consistent documentation/package quality. Docs and packages can be patchy and inconsistent, or incomplete.