

Excel Mastery Project and Certification

(70 Marks)

First Name: **ABIMBOLA**Last Name: **AKERELE**State Code: **BN/23C/0217**

Mentorship Mode: Full Short Term

Instruction:

You are required to answer at-least 4 Questions in all, Number **one is compulsory**.

Question 1.

You are provided with a link which contain Dataset of 19,000 online job posts from 2004 to 2015. Use the link below to download the dataset.

<https://www.kaggle.com/datasets/madhab/jobposts>

About Dataset

Job Posts dataset

The dataset consists of 19,000 job postings that were posted through the Armenian human resource portal CareerCenter. The data was extracted from the Yahoo! mailing group <https://groups.yahoo.com/neo/groups/careercenter-am>. This was the only online human resource portal in the early 2000s.

A job posting usually has some structure, although some fields of the posting are not necessarily filled out by the client (poster). The data was cleaned by removing posts that were not job related or had no structure.

The data consists of job posts from 2004-2015

Answer the following:

- a. Carry out a complete Data Cleaning Process in the Dataset such as removing null or empty values, data duplicates, missing values if any

- b. Using Data Validation to validate your work, ensure to write down each validation process used
- c. Prepare a KPI, Charts Visuals Dashboard of not less than four and more than 6 Charts showing the following
 - Different job titles
 - Total Number of Companies
 - Average Salary Range
 - Total Number of Job Posts
 - Job Post by Locations
 - Job Duration for Each Job title
 - Other Vital Information

Note: Line charts should be used to show data distribution, Bar-charts and column charts for categorical data, pie chart for categorical data not more than 4 categories

- d. From question c above, what can you deduce as report of the importance
- e. of online jobs and the advice you can give to the massive to maximize online job posts ?.
- f. State two uses each of the following charts
 - **Line Charts**
 - 1. It represents the change in a quantity(value) with respect to another quantity(value).
 - 2. It is used in monitoring behavior in a set of data by showing trends in the series
 - **Bar charts**
 - 1. are used to show the comparisons or distribution of between different variables
 - 2.
 - **Pie Charts**
 - 1. It is helpful for showing the relationship of parts of dataset to the whole dataset
 - 2. It shows the percentage of each data points in the data set
 - **Combo Charts**
 - 1. It combines two or more chart types in a single chart
 - 2. it allows you to compare two variables that have different magnitudes and scales of measurement.
 - **Column Charts**

1. for comparing data across categories
2. it is used to visualize changes in values over time and identify trends or patterns.

Question 2:

Explain the following in data analysis:

- a. Data Validation
- b. Big Data
- c. Data Consolidation
- d. Data Modelling
- e. Functions in Excel
- f. Data Wrangling/Cleaning
- g. Conditional Formatting

Data validation refers to the process of ensuring the accuracy and quality of data by verifying and validating data that is collected before it is used. Data validation in Excel is a feature that allows you to control the type and format of data entered into a cell. It helps ensure data accuracy by setting rules or criteria for the input.

Procedure for data validation:

- Select a cell/ range to apply data validation.
- Go to the "Data" tab, then click on the "Data Validation" option.
- In the Data Validation dialog box, in the **allow drop down** specify criteria to fit in the option of your choice(e.g whole number, list, date, time, text length and others) and set to exact specifications.
- Input Message: You can add a message that appears when the cell is selected, informing the user on what input is allowed.
- Error Alert: Create an alert to notify users if they enter invalid data. This can be a warning or an error message.
- Once set, the data validation rules will be applied to the selected cells.

The primary goal of data validation is to improved data quality and consistency, prevent or reduce data entry errors, aid faster data processing, improved decision-making

Big Data

Big data is a combination of structured, semi structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.

Structured data is highly organized and fits neatly into traditional databases. Semi-structured data, like JSON or XML, is partially organized, while unstructured data, such as text or multimedia, lacks a predefined structure.

Some examples are transaction processing systems, customer databases, documents, emails, medical records, internet clickstream logs, mobile apps and social networks. Big data is a collection of data from many different sources and is often describe by five characteristics: volume, value, variety, velocity, and veracity.

Companies use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and take other actions that, ultimately, can increase revenue and profits. Businesses that use it effectively hold a potential competitive advantage over those that don't because they're able to make faster and more informed business decisions.

Big data is also used by medical researchers to identify disease signs and risk factors and by doctors to help diagnose illnesses and medical conditions in patients. In addition, a combination of data from electronic health records, social media sites, the web and other sources gives healthcare organizations and government agencies up-to-date information on infectious disease threats or outbreaks.

The challenges of big data include difficulty in making the data accessible to data scientists and analysts, especially in distributed environments that include a mix of different platforms and data stores. The process of integrating sets of big data is often also complicated.

Data consolidation is the process of combining data from multiple sources, cleaning and verifying it by removing errors and storing it in a single location like data warehouse.

ETL is a important Part of data consolation. ETL means Extract data from the different source, Transform the data so that it is better suited for analytics (sorting, cleaning and removing duplicate), Load the data into the target database

It makes all data management information available quickly and easily, and having all data in one place increases productivity and efficiency. Consolidation also reduces operational costs and facilitates compliance with data laws and regulations. The main benefit is that it allows you to analyse your data at a later stage to make decisions based on facts and data.

Data modeling is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures.

Data modeling is the process of producing a diagram of relationships between various types of information that are to be stored in a database that helps us to think systematically about the key data points to be stored and retrieved, and how they should be grouped and related.

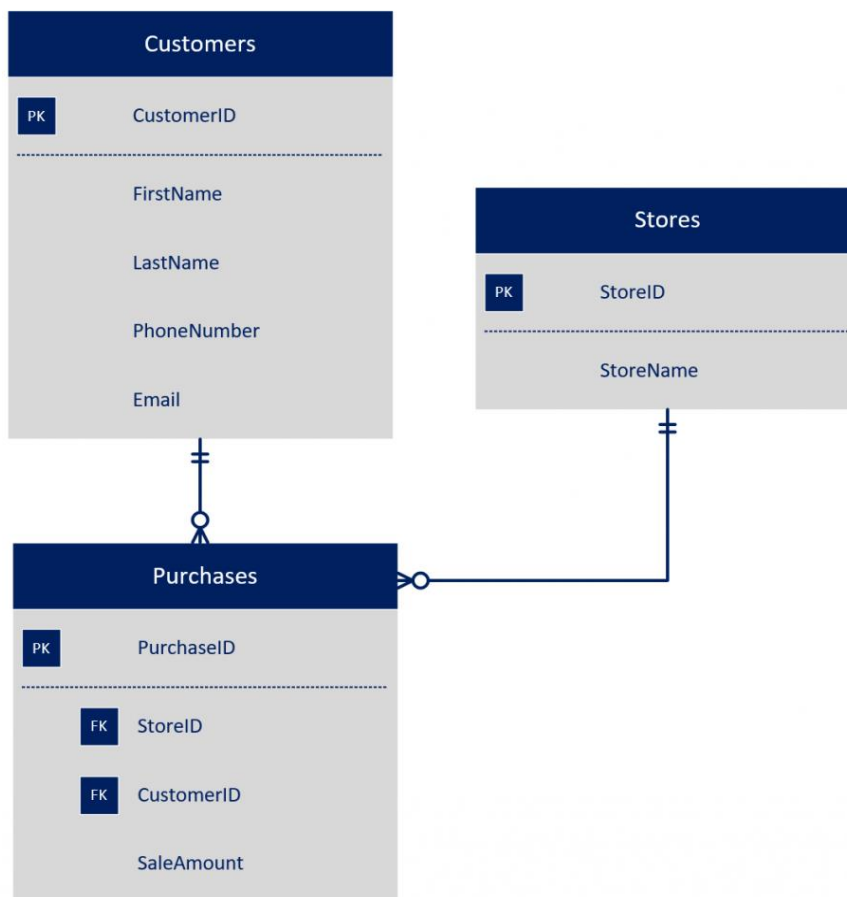
Types of relationships in data modelling

One-to-one: When only a single instance of an entity is associated with the relationship, it is termed as '1:1'.

One-to-many: When more than one instance of an entity is related and linked with a relationship, it is termed as '1:N'.

Many-to-one: When more than one instance of an entity is linked with the relationship, it is termed as 'N:1'.

Many-to-many: When more than one instance of an entity on the left and more than one instance of an entity on the right can be linked with the relationship, then it is termed as N:N relationship.



Functions in Excel

A function is a predefined formula that performs calculations using specific values in a particular order.

In order to work correctly, a function must be written a specific way, which is called the syntax. The basic syntax for a function is an equals sign (=), the function name (SUM, for example), and one or more arguments.

Examples of excel functions include sum, sumif, sumifs, count, counta, ifs, if, concatenate, large, min, rank, vlookup, xlookup, rate, workdays, date, trim, clean, upper, lower and many more.

Data cleaning is the process of making sure that the data is accurate and consistent. Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. This process is aimed at improving the quality and reliability of your data, making it more accurate and useful for analysis. Examples include finding missing values, replace values, data formatting

Data wrangling process also known as a **data munging process**, consists of reorganizing, transforming and mapping data from one "raw" form into another in order to make it more usable and valuable for a variety of downstream uses including analytics.

Data wrangling can be defined as the process of cleaning, organizing, and transforming raw data into the desired format for analysts to use for prompt decision-making. It enables businesses to tackle more complex data in less time, produce more accurate results, and make better decisions. The exact methods vary from project to project depending upon your data and the goal you are trying to achieve.

- Data wrangling helps to improve data usability as it converts data into a compatible format for the end system.
- It helps to quickly build data flows within an intuitive user interface and easily schedule and automate the data-flow process.
- Integrates various types of information and their sources (like databases, web services, files, etc.)
- Help users to process very large volumes of data easily and easily share data-flow techniques.

Tools used for these process include Excel Power Query, Tabula, Google DataPrep , Data wrangler

Conditional formatting allows you to dynamically change a cell's text style and background color based on custom rules you set. It is a feature to automatically change the font properties of a specific cell, row, column.

Steps to conditional formatting:

1. Select the range you want to format.

2. Go to Format in the top menu. Select Conditional formatting.
3. Select either Single color or Color scale.
4. Apply a formatting rule (or rules) and a formatting style.
5. Click Done

Question 3.

- a. Write out 30 Excels functions and state each of them uses with practical examples (attach just screenshots)
- b. Write out 45 Excel Short Cut Keys and state their uses

Question 4.

- a. Prepare a simple student grading system and compute the following using the **IF** or **Vlookup** Statement to validate the score grade base on students score. Append the functions for each cell. See criteria for computation below

Score	Grade	remark
0 - 30	F	FAIL
31 - 40	E	POOR
41 – 49	D	FAIR
50 – 60	C	AVERAGE
61 – 70	BC	GOOD
71 - 74	B	V. GOOD
75 - 100	A	EXCELLENT

IF FUNCTION

=IF(B5>=75,"A", IF(B5>=71,"B", IF(B5>=61, "BC", IF(B5>=50,"C", IF(B5>=41, "D", IF(B5>=31, "E",IF(B5<=30, "F"))))))))

VLOOKUP

=VLOOKUP(C4,\$H\$5:\$I\$11, 2,FALSE)

- b. State the benefits of tables in data analysis in Excel
- **sorting and filtering your data quickly**
 - **automatic update of charts and Pivot Tables**
 - **it makes it easier to validate data**
 - **re-naming a range**
- c. What is Pivot table used for
- A PivotTable is a tool used to calculate, summarize, and analyze data**
- d. Show a simple chart using slicer for a demo data

Question 5.

- a. Write out 10 Excel Add-ins and state their uses
1. **Solver** helps to define and solve optimisation problems. It helps to find the optimal value in a cell, called the target cell, on any Excel spreadsheet. It can be create Solver models to perform common data analytical tasks
 2. **The Analysis ToolPack** is a Microsoft add-in that allows statistical analysis such as correlation analysis, descriptive statistics or histograms.
 3. **Power query** is made to help access and explore data in Excel like with Business Intelligence tools, by allowing to import, transform or combine multiple data sources.
 4. **Risk analysis** add-in uses Monte Carlo simulation to helps to make better decisions through risk modeling and analysis.
 5. **Mini Calendar and Date Picker** just pick a date in a calendar and it will be automatically formatted as date.
 6. **Geographic Heat Map** helps to organize parts of the world using different colors so you'll have a better understanding of where everything is spatially.
 7. **Excellent Analytics** add-in to import your Google Analytics marketing data into a spreadsheet.
 8. **PowerApps** helps to connect the dots between data points and gives the ability to share these findings with your team.
 9. **FRED** significantly cut down your research time when hunting down economic information quickly and efficiently
 10. **Bookmarks** create a place to keep all of your favorite directories and workbooks.
- b. State two uses of excel Macros
1. **It is used to automate task thereby saving time and efforts**

2. It is used for organizing data in an Excel spreadsheet

- c. Carry out a simple linear regression in excel

Question 6.

- a. Create a simple VB form to receive the following inputs

- First Name
- Last Name
- Gender
- Email
- Username
- Password
- Course of Study

With a submit button

- b. Write Simple VB to do the following

- Sum two numbers together
- Find the average of four numbers

15 Marks

MODE OF SUBMISSION

Create a Github account with your name and also create a public repository called [Mini Excel Project](#). Upload your project to you Github and share link to the assignment group. Your File should carry the following details

- First Name
- Last Name
- State Code / ID
- Mentorship Mode (Full or Short Term)

NOTE: Deadline for Submission is 16th March, 2024

Defense – 17th March, 2024

Time of Defense 8:00PM

Venue: Google Meet

Mode of defense: You will be sharing your screen, Endeavor to Charge your laptop during the day of defence