**README Data Analysis PROJECTS MeriSKILL**

**After Completing any of the two projects** [**click here to claim your certificates**](https://docs.google.com/forms/d/e/1FAIpQLSdUYKNNxmr1ecyYq6N9boOqOGmfGB3rhvyvA1d1caYdMleRYA/viewform) **or Visit https://www.meriskill.com/certificate**

**Title of the project 1: “**Sales Data Analysis”

[Guide - Click Here](https://drive.google.com/file/d/1rMSUuu9Zcv9lxHdbKoTcYAsXZSjCHuA3/view?usp=drive_link)

**Purpose:**Analyze sales data to identify trends, top-selling products, and revenue metrics for business decision-making.

**Description:** In this project, you will dive into a large sales dataset to extract valuable insights. You will explore sales trends over time, identify the best-selling products, calculate revenue metrics such as total sales and profit margins, and create visualizations to present your findings effectively. This project showcases your ability to manipulate and derive insights from large datasets, enabling you to make data-driven recommendations for optimizing sales strategies.

**Title of the project 2: “**Diabetes Patients”

[Guide - Click Here](https://drive.google.com/file/d/1rMSUuu9Zcv9lxHdbKoTcYAsXZSjCHuA3/view?usp=drive_link)

**About Dataset**

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney  
Diseases. The objective of the dataset is to diagnostically predict whether a patient has diabetes  
based on certain diagnostic measurements included in the dataset. Several constraints were placed  
on the selection of these instances from a larger database. In particular, all patients here are females  
at least 21 years old of Pima Indian heritage.2  
From the data set in the (.csv) File We can find several variables, some of them are independent  
(several medical predictor variables) and only one target dependent variable (Outcome).

**Title of the project 3:**“HR Analytics"

[Guide - Click Here](https://drive.google.com/file/d/1rMSUuu9Zcv9lxHdbKoTcYAsXZSjCHuA3/view?usp=drive_link)

**Tasks to perform:**

**Data Cleaning:**

Deleting redundant columns.  
Renaming the columns.  
Dropping duplicates.  
Cleaning individual columns.  
Remove the NaN values from the dataset  
Check for some more Transformations

**Data Visualization:**

∙        Plot a correlation map for all numeric variables

∙        Overtime

∙        Marital Status

∙        Job Role

∙        Gender

∙        Education Field

∙        Department

∙        Business Travel

∙        Relation between Overtime and Age

∙        Total Working Years

∙        Education Level

∙        Number of Companies Worked

∙        Distance from Home