بسم الله الرحمن الرحيم

Excel for Data Analysis – Full Course for Beginners

Project One

**Salaries of Data Science Jobs**

**(Jobs Calculator / Dashboard)**

**Content:**

* Overview.
* Part 1: Prepare Data for Analysis.
* Part 2: Analyze the Data.

**Written by:**

**Abobaker Ahmed Khidir Hassan**

Overview

About the Project

This course is an Excel course for Data Analysis. It’s starting from installing Excel up to creating a full project using DAX.

This project is the first full project in the course and it use the source data contains 16 columns and 32,762 rows about the Data Science Jobs Posts to create a calculator or mini dashboard that helps taking decision which job title must be focused in.

Lessons Have Been Taken:

* What is Excel?
* Formulas Intro
* Function Intro
* Logical Functions
* Math Functions
* Statistical Functions
* Array Formulas
* Lookup Function
* Text Functions
* Date and Time Functions
* Charts Intro
* Charts Advanced
* Charts Statistics
* Sparklines
* Tables
* Formatting
* Collaboration
* Project #1: Build Dashboard
* Project #1: Share Projects

Data

The original data is consists of 16 columns and 32,762 rows/records.

| **Column Name** | **Description** | **Inferred Data Type** |
| --- | --- | --- |
| job\_title\_short | Shortened version of job title | string |
| job\_title | Full job title and department | string |
| job\_location | City and state of the job | string |
| job\_via | Platform where the job was posted | string |
| job\_schedule\_type | Full-time, part-time, etc. | string or categorical |
| job\_work\_from\_home | Whether the job allows remote work | boolean |
| search\_location | Location used for searching jobs (possibly user input) | string |
| job\_posted\_date | Date since the job was posted (appears to be in float days format) | float or int |
| job\_no\_degree\_mention | Whether the job mentions that a degree is *not* required | boolean |
| job\_health\_insurance | Whether health insurance is offered | boolean |
| job\_country | Country where the job is located | string |
| salary\_rate | Salary time unit (e.g., year, hour) | string (categorical) |
| salary\_year\_avg | Average annual salary | float |
| salary\_hour\_avg | Average hourly salary (may be null) | float or null |
| company\_name | Employer offering the job | string |
| job\_skills | List of required skills | list[string] or string |

Preparing Data for Analysis

In this section, I will clean and applying many concepts about the data preprocessing as what we will se.

Step 1: Create a Copy of the Data

This is my first step before doing anything, and we know how it’s important to save the original data in a safe place.

I named the new worksheet WorkSpace.

Step 2: Formatting Data as a Table

To make referencing more efficient and reduce the referencing error those maybe happen in the cell referencing.

I named it jobs.

Step 3: Formatting

I set the format/data types of the fields then removing spalling errors, blanks to make sure the analysis will be applied in a clean data.

Note:

I note that there is some uncompleted data in some columns such as:

job\_location, job\_via, job\_schedule\_type, and job\_skills.

And I note 10636 of the salaries are yearly salaries, and the others (22036) are hourly salaries.

All of these will be in mind during the analysis.