# AN INTRODUCTION TO DATA ANALYSIS WITH MICROSOFT POWER BI

March 28, 2021


## Abstract

Enclosed are notes captured while completing Power BI training during March 2021.



Timothy D. Vines

VinesTimm@Gmail.Com

# Table of Contents

Introduction to Data Analysis with Microsoft Power BI 2

1. Preparing data for analysis 2
2. Creating key metrics 2
3. Comparison analysis 2
4. Trend Analysis 2
5. Ranking Analysis 2
6. Interactive Dashboards 2
7. Contribution Analysis 3
8. Variance Analysis 3
9. Correlation Analysis 3
10. Frequency Analysis 3
11. Summary Tables 4

**Certification Earned**: 5

# Introduction to Data Analysis with Microsoft Power BI

Certification Earned: https://drive.google.com/file/d/1XgMkngiH35eeqpvYJCNrQS2cdG1WNT0g/view?usp=sharing
Training video: https://lendmarkfinancial.udemy.com/course/introduction-to-data-analysis-with-microsoft-power-bi/learn/lecture/17879838#overview

**Notes:**

1. Preparing data for analysis - Rules:
    a. Each Column is a field of data
    b. Name the columns in a meaningful way
    c. Each row is a record of data
    d. Do not include subtotal and grand totals in your data - Only data
    e. Get the data types correct (*Text, numeric, date*)

2. Creating key metrics (*Use **Cards** / Tables to show KPI: Important method to understand the health of the business*) - Examples:
    a. Totals - What is the total sales / profit?
    b. Average - What is the Average Sale amount per transaction?
    c. Count - How many transactions have been processed?
    d. Ratio - The proportion of one value to another
    e. Highest - What is the highest individual value
    f. Lowest - What is the lowest individual value

3. Comparison analysis (*Compare the value of one item to another* - **Column and Bar Graphs** are used to easily show comparison):
    a. Understand the highest to lowest values and by how much
    b. Visualizing data to make analysis easier - Column and Bar Graphs
    c. Create Themes to default text size and color

4. Trend Analysis (*Easily understand how data changes over time* - **Line and Area Graphs**):
    a. Answer questions: Is data moving up, down stable or volatile
    b. Understand how data changes over years, quarters, months, days, hours, etc
    c. Use line and area graphs
    d. Use of Trendlines

5. Ranking Analysis (*Sort with Table Chart and use of **Ribbon Chart** to show ranking over time*)
    a. Simplest form of ranking analysis is to **sort**
    b. Perform top / bottom 10 analysis using the **top** filter for a dimension based on a measure

c. Calculate rankings - Using the **RankX** / **All** functions in Dax to rank standings based on a dimension

6. Interactive Dashboards - Interactive reports combining:
    a. KPIs
    b. Using Comparison Analysis
    c. Trend analysis
    d. Ranking Analysis

7. Contribution Analysis - Percentages:
    a. Use **Pie Graph** to display percentages
    b. Can change measures to show percentage of grand total, row, or column when using a **matrix** chart (*Basically a pivot table*)
    c. **100% Stacked Colum/Bar chart** can be used to show the contribution of an item towards 100% of another item - Turn data labels on to see the percentages on the graph

8. Variance Analysis - **Running Total** / **Actual vs Budget** / **Time Calculations** - Shows how different data is from other data. Applications include:
    a. Understanding the difference between actual and budget
    b. Calculate the difference from the previous month or year - Calculate Previous Month's data:
        A. Sales Prev Month = **CALCULATE**([Total Sales],**PREVIOUSMONTH**(Data1[Order Date].[Date]))
        B. Sales Prev **Year/Month** = CALCULATE([Total Sales],**DATEADD**(Dates[Master Date],-1,YEAR))
        C. Total YTD = **TOTALYTD**([Total Sales],Dates[Master Date])
        D. Diff From PM Sales = [Total Sales] - [Sales Prev Month]
    c. Use a date master file for time intelligence calculations

9. Correlation Analysis - Relationship between two variables - **Scatter Chart** (*EX: Relationship between Sales ang Profit*):
    a. Can add a date (month) to the **Play Axis** to play the correlation over time and slice on the year

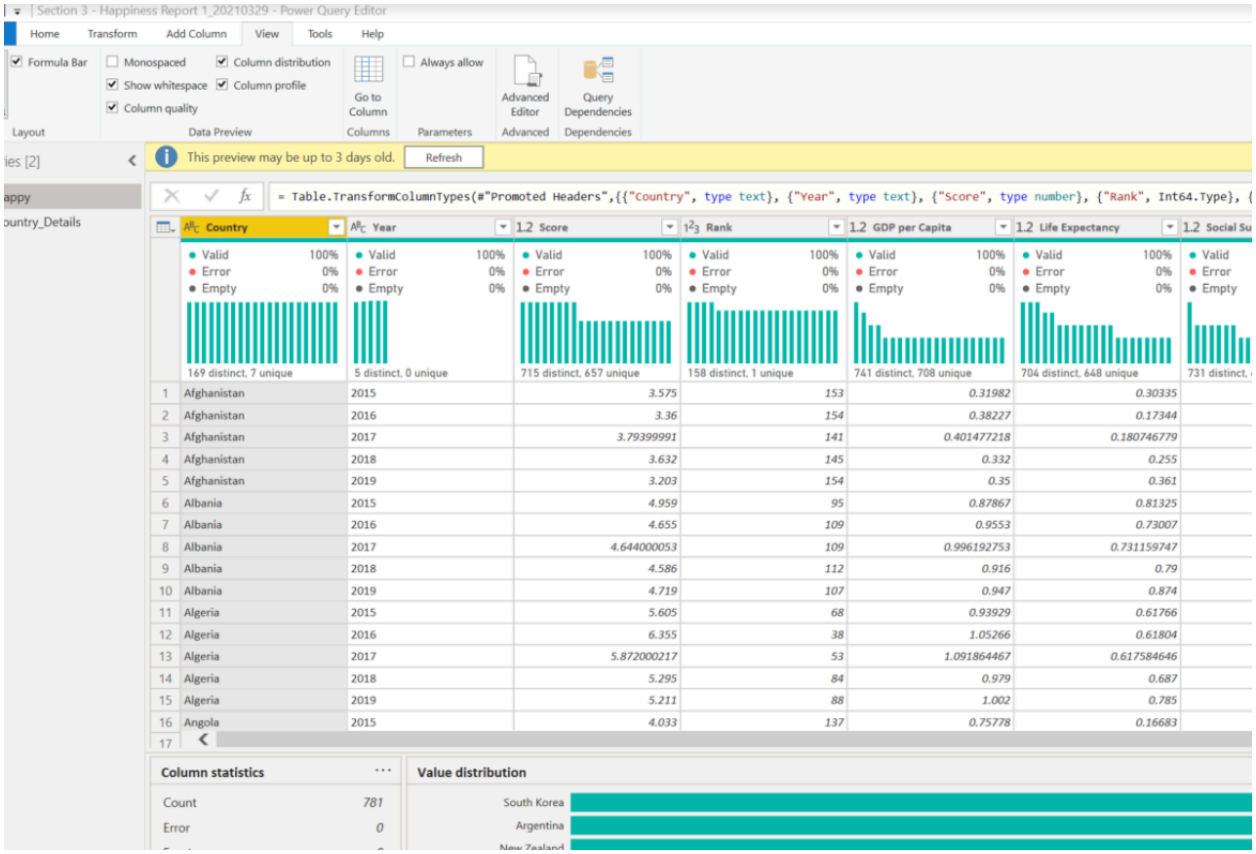
10. Frequency Analysis - **Bucketization**/Groupings of data:
    a. Number of occurrences within a grouping of data (*EX: Human Resources analyze ages 20 to 30, 30 to 40, etc*) - Two ways to accomplish:
        A. Can create a group on the dimension with bins, or
        B. Use Switch Formula to create a new column:
            1. Go to the data view
            2. Click New Column

3. Age Group (Switch Formula) = **SWITCH**(True(), Master[Age]>50,"4. >50",Master[Age]>40,"3. 40 to 50",Master[Age]>30,"2. 30 to 40","1. 20 to 30")
4. Salary Switch = **SWITCH**(True(),Master[Salary Monthly]> 4000, "e. >4000",Master[Salary Monthly]> 3000, "d. 3000 to 4000", Master[Salary Monthly]> 2000, "c. 2000 to 3000",Master[Salary Monthly]> 1000, "b. 1000 to 2000","a. 0 to 1000")

b. Use Column Graph to create histograms - Use Stacked Column Graph to show the frequency

11. Summary Tables - Working with Large Data sets:
    a. Create calculations to create summarized values - Create a summary table on the Modeling tab to reduce the calculation time required:
        A. Country = SUMMARIZE(Data1,Data1[Country],"Total Sales - Summary Table", sum(Data1[Sales]), "Total Profit - Summary Table",Sum(Data1[Profit]),"Avg Sales - Summary Table",AVERAGE(Data1[Sales]))

    b. Create relationships between tables and calculations amongst those tables:
        A. Once relationship is created, can calculate with fields from table to table:
        B. Sales 2013 = **CALCULATE**(Sum(Data1[Sales]),Data1[Year]=2013)
        C. #Trans2013 = **CALCULATE**(DISTINCTCOUNT(Data1[Order ID]),Data1[Year]=2013)
        D. Cardinality and Cross Filter Direction - Cross Filter means we can filter in either one direction or with both directions
        E. Consolidating tables from the query editor use the **Merge** option:
        F. Can got to the Query Editor then View Column Quality/Distribution/Profile to see data information




## Certification Earned:

https://drive.google.com/file/d/1XgMkngiH35eeqpvYJCNrQS2cdG1WNT0g/view?usp=sharing


Certificate of Completion
This is to certify that Timothy D Vines successfully completed 5 total hours of Introduction to Data Analysis with Microsoft Power BI online course on March 31, 2021
Ian Littlejohn
Ian Littlejohn, Instructor
&
Udemy
Certificate no: UC-0bb5b95b-6c17-4c79-b30f-010283892741
Certificate url: ude.my/UC-0bb5b95b-6c17-4c79-b30f-010283892741
Version 3
#BeAble