## horizontal line



# Methodology to perform “VTAC interview and task” project

16.05.2024

**─**

Ali Keyhani

East Melbourne,

# My Methodology

Completing this task involved **four steps** for me:

## 1- Preparing and Cleaning Data (Data Preprocessing)

Although the given dataset was quite small and could be easily managed using Excel, I preferred to use the Python programming language to showcase my skills in this area.

Initially, I merged two datasets and addressed issues such as:

• Duplicate data

• Incorrect data types, especially in the date fields

• Logical data errors, such as when the number of paid\_apps was recorded as being higher than all\_apps.

Eventually, I created a clean and merged dataset for use in data visualization.

## 2- Analyzing insights that can be extracted from this data table and selecting the most appropriate visuals for the best representation.

Given the task defined for me (to compare cumulative applications in the intake periods), I tried to extract the most relevant insights from the raw data.

## 3- Designing and implementing data models, Measures, Calculated columns, and Calculated tables based on selected visuals and insights.

### dim\_date Table In this phase, I decided to use a dimension table named dim\_date in my data model to optimally implement my visuals.

In this table, I created a field named “Financial Year” to divide the dataset into two groups, 2021 and 2022.

Also, to ensure that the months of each time period from August (month 8) to January (month 1) were correctly sorted in the tables and charts, I created a field named Month Counter, in which August was set as 1, January as 6, up to July, which was set as 12.

Here is the DAX code I used to create the dim\_date table:

dim\_date =

VAR StartDate = MIN('Fact\_apps'[date\_applied])

VAR EndDate = MAX('Fact\_apps'[date\_applied])

VAR Dates = CALENDAR(StartDate, EndDate)

RETURN

ADDCOLUMNS(

Dates,

"Year", YEAR([Date]),

"Month", MONTH([Date]),

"MonthName", FORMAT([Date], "MMMM"),

"Quarter", "Q" & FORMAT(QUARTER([Date]), "0"),

"Day", DAY([Date]),

"Weekday", WEEKDAY([Date]),

"WeekdayName", FORMAT([Date], "dddd"),

"IsWeekday", IF(WEEKDAY([Date], 2) <= 5, "Weekday", "Weekend"),

"YearMonth", FORMAT([Date], "YYYY-MM"),

"Financial Year",

IF(

MONTH([Date]) >= 8,

YEAR([Date]),

YEAR([Date]) - 1

),

"Month Counter",

SWITCH(

MONTH([Date]),

8, 1, // August

9, 2, // September

10, 3, // October

11, 4, // November

12, 5, // December

1, 6, // January

2, 7, // February

3, 8, // March

4, 9, // April

5, 10, // May

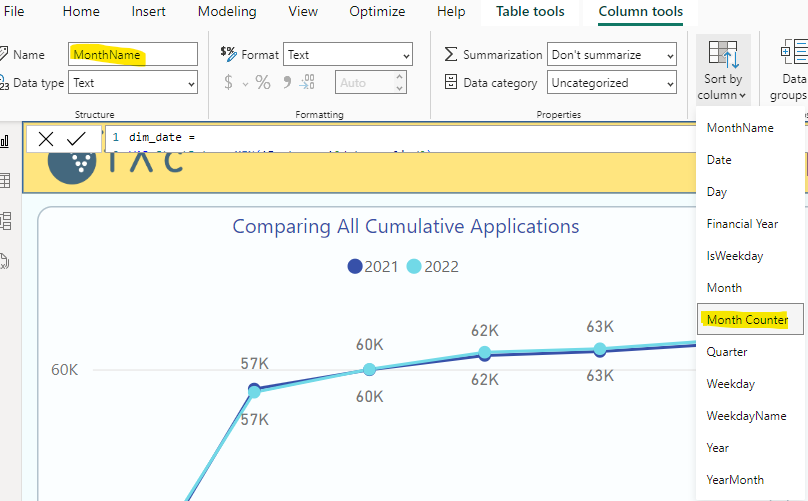
6, 11, // June

7, 12 // July

)

)

In the next step, I sorted the MonthName column using the Month Counter.



### Measures

Additionally, to cumulate the data, I created a Measure Table and developed all the desired measures within it. Some of these measures include:

all\_apps\_ = sum(Fact\_apps[all\_apps])

all\_apps\_cumulative =

CALCULATE(

'measure\_table'[all\_apps\_],

FILTER(

ALL('dim\_date'),

'dim\_date'[Financial Year] = SELECTEDVALUE('dim\_date'[Financial Year]) &&

'dim\_date'[Date] <= MAX('dim\_date'[Date])

)

)

all\_apps\_cumulative\_varicance =

var t1 = ([all\_apps\_cumulative] - [all\_apps\_cumulative\_PreviousYear] ) / [all\_apps\_cumulative\_PreviousYear]

return

IF([all\_apps\_cumulative\_PreviousYear] <= 0 && [all\_apps\_cumulative] > 0 , "-", t1)

## 4- Implementing and Formatting the Visuals

Finally, I placed the visuals on the page and formatted them according to the color scheme of the vtac.edu.au website.

# Overview of Visualizations

* **Comparing All Cumulative Applications:** A line chart showing the cumulative applications over time for both years.
* **Comparing Paid Cumulative Applications:** A bar chart comparing the cumulative paid applications by month for both years.
* **Comparing Unpaid Cumulative Applications:** A bar chart comparing the cumulative unpaid applications by month for both years.
* **Applications Table:** Detailed breakdown of all applications.
* **Monthly Breakdown of Paid and Unpaid Applications:** A stacked column chart showing the percentage of paid and unpaid applications for each month, comparing the financial years 2021 and 2022.

# Key Insights

The exact match of total applications received in 2021 with those in 2022 suggests possible data manipulation.

Additionally, the percentage of unpaid apps in Januarys is also a matter of consideration.