Health Care Data – *Requirement Gathering*

Requirement Gathering

**Identify Stakeholders**

* Determine the primary stakeholder and establish a point of contact who might be the domain experts or leaders who will eventually use the dashboard.

**Understand Business Objectives**

* Through meetings and calls with stakeholders you should get an outline of Goals from the entire endeavor. Asking open-ended questions will help you gain more insights to understand the data and how this dashboard will help achieve a specific business goal.

**High Level Data Study**

* A high-level overview of data is required before you initiate discussions around scope, metrics and other granular topics. So here we will try to understand the data in terms of:
  + Data Sources
  + Column Description
  + Data Type
  + Volume and Frequency
  + Data Quality – Missing Values or Anomalies

**Define Scope**

* This is the perfect stage to discuss Key Metrics, KPIs & Deployment Timelines. Document the calculations, time frames & scope which will help in setting the expectations and avoiding any future disagreements. Also, as best practice, remember to keep a 20% buffer while finalizing deadlines because it’s always better to over-deliver after a standard commitment than to underdeliver after an extraordinary delivery pitch.

**What is the problem statement for this exercise?**

* **Overall** **Objective**
  + Track current status of patient waiting list
  + Analyze historical monthly trend of waiting list in Inpatient & Outpatient categories
  + Detailed specialty level & age profile analysis
* **Data** **Scope**
  + 2018 – 2021
* **Metrics**
  + Average & Median Waiting List
  + Current Total Wait List
* **View** 
  + Summary Page
  + Detailed Page for Granular Analysis

Data Collection

There are over 200 different data connectors available to collect data inside power bi, but below are the most widely used connectors:

* Excel/CSV
* Folder Connection
* SQL Server or Any Database
* Power BI services
* Cloud Platform – Azure, AWS, GCP etc.

This is the stage where you decide the source of your data. This step is very important because this will also define how you are going to refresh the dashboard after deploying the solution. For today’s example we are going to use a central folder which will host all the files required for the dashboard refresh process.

I’m going to download the zip file from the above download button and extract the file inside a new folder on my desktop. This will act as my data repository going forward.

Data Transformation

**Data transformation** is a process of changing the structure of your data or applying additional steps which will clean or process your data for final usage. We do these transformations in the Power Query Editor which is inbuilt into Power BI.

**Renaming Columns**

* While studying the data, I noticed that the Specialty column in Inpatient data is named as Specialty\_Name, whereas in Outpatient it is named as just Specialty. So, we will rename the Outpatient column to match with the Inpatient data. Make sure to name it the same, otherwise it will create an issue in the following steps.

**Rearranging Columns**

* We will now rearrange our columns of the outpatient so that it matches with inpatient. You can just left click on the column header and drag it to the required position. Now while rearranging, I noticed that we have an additional column in Inpatient i.e. Case\_Type which is missing from Outpatient. So, let’s create one additional column in Outpatient table called Case\_Type by:
  + Go to Add Columns
  + Click on Custom Columns
  + Name the column as Case\_Type
  + Enter the formula =”Outpatient”
  + Remember to place this new column at the same position as inpatients

**Appending Tables**

* Now that both our tables have the same column structure, we can safely append them together. To do that come to Home tab and click on “**Append Queries**” button and click on “**Append Queries as New**”. Select the 1st table as inpatient and 2nd table as outpatient. Rename this new table as “All\_Data”. Finally click on “Close & Apply”.

**Appending Tables (Continued)**

* Observe that Age\_Profile & Time\_Band columns have some redundant data, so first use the Replace function button in power query to clean the data, for e.g.: “18+ months” and “18 month +”, both are the same, so replace either one to match the other. Secondly there are some trailing blanks in values of these columns so remove trailing blanks by using the Trim function button.

**Data Modeling**

* Data modelling is a way to create relationships with one table to another, so that we can fetch valuable information from them in our reporting layer.
* Let’s jump into the **Data Modelling View**, which is located at the left-hand panel on Power BI. We will be using All\_Data from now onwards, so we can safely hide inpatient and outpatient data. We can also stop it from loading into the data model by disabling it from the power query editor. Just right click on the table name in power query and uncheck Enable Load.
* Now since specialty name is one of the key attributes that we are looking at in this project, let’s focus on that column now. As you have seen from the data, we have a huge number of specialties available and using all of them in our report layer directly will create a clutter in our visualization. A better approach would be to distribute them in buckets. So, to do this I have created a specialty mapping file which you will find in the downloaded resources. Let’s import that file in power bi to create the relationship with All\_Data.
* Once you import this file, Power BI should auto detect relationship and connect both the tables. However, if it does not then you can do it manually by following the steps below:
  + Go to the model view
  + Click on Specialty\_Name column in All\_Data
  + Drag this column over the Specialty\_Name column in Mapping table
* Now you should see a line connecting both the tables and an arrow pointing towards All\_Data from the Mapping table. This means that we have created a relationship between both the tables. The arrow signifies the filter context and tells you that now you can use columns from the Mapping table to filter data in All\_Data.

Visualization Blueprint

For this exercise, I have already created a blueprint, however in live scenarios you will sit down with your team and create a wireframe of the required dashboard. You will then get this approved from the end stakeholder before starting your development activities.

Dashboard Layout & Design

Before you start the designing process, I usually enable below two properties from the View tab. This helps me place and align my visuals in a uniform manner.

* Gridlines
* Snap to Grid

Now let’s use DAX to create our measures which will be used in our visuals. For now, we will create 2 measures for calculating the Latest Month & Previous Year Wait List

Latest Month Wait List = CALCULATE(SUM(All\_Data[Total]),All\_Data[Archive\_Date] = MAX(All\_Data[Archive\_Date])) + 0

PY Latest Month Wait List = CALCULATE(SUM(All\_Data[Total]),All\_Data[Archive\_Date]= EDATE(MAX(All\_Data[Archive\_Date]),-12)) + 0

Now as per our design blueprint, insert these 2 measures in a card visual. After this we will create a blank table where we will store calculation method headers, i.e. in our dashboard we want to show Average values and Median values.

So, enter a blank table from power bi and enter 2 row items manually i.e. Average & Median. Now with the newly created table create a button slicer, so that users can toggle between the two calculations.

Now create measures below which will help us get the calculation we need and to make a few chart titles dynamic

Median Wait List = MEDIAN(All\_Data[Total])

Average Wait List = AVERAGE(All\_Data[Total])

Avg/Med Wait List = SWITCH(VALUES('Calculation Method'[Calc Method]),"Average",[Average Wait List],"Median",[Median Wait List])

Dynamic Title = SWITCH(VALUES('Calculation Method'[Calc Method]),"Average","Key Indicators - Patient Wait List (Average)","Median","Key Indicators - Patient Wait List (Median)")

NoDataLeft = IF(ISBLANK(CALCULATE(SUM(All\_Data[Total]),All\_Data[Case\_Type]<>"Outpatient")),"No data for selected criteria","")

NoDataRight = IF(ISBLANK(CALCULATE(SUM(All\_Data[Total]),All\_Data[Case\_Type]="Outpatient")),"No data for selected criteria","")

**Summary Page**

* Now place the charts based on our blueprint i.e. doughnut, clustered column chart & top five Multi Row cards. And remember to use the new measure which is Avg/Med Wait List in the values section.
* Finally in the line chart at the bottom use Total column directly along with the Archive\_Date. Remember to add a visual filter for Case\_Type. So, one chart will show Day Case & Inpatients, and the other chart will show Outpatients. Add slicers for Archive\_Date, Case\_Type and Specialty.

**Detailed View**

* Add a new page here and add a matrix view using the Archive\_Date, Specialty\_Name, Age\_Profile, Time\_Bands, Case\_Type and Total.

**Tooltip Page**

* Create a new page that will be used as a tooltip. Add a chart to show the Specialty and Total waitlist. Also, add a card to show the Total sum of Wait List. Now set this page as a tooltip by going to formatting >> Page Information >> Enable Allow Use as Tooltip.
* Now go back to the summary page and select the line chart. In the General section of formatting, go to Tooltips and select the page i.e., the new tooltip page.

Adding Interactivity

Now add interactivity in your dashboard like navigation buttons, chart alt display text and hovering info.