**PwC Switzerland Data Analytics PowerBI Virtual Internship**

**Table of Contents:**

**• Flow of work-:**

Step 1- Upload Data

Step 2-Cleaning data

Step 3-Transform data

Step 4-Load data

**• Data Preparation**

Data Modelling

Writing DAX

**• Data Visualization**

**• Data Analysis**

**• Insights**

**Task 2-Call Centre Trends**

Visualizing customer and agent behaviour. Create a dashboard in Power BI for Call Centre Manager that reflects all relevant Key Performance Indicators (KPIs) and metrics in the dataset.

**•Problem Statement**

The purpose of this analysis is to create a dashboard in PowerBI for call canter manager that reflects all relevant Key Performance Indicators (KPIs) and metrics in the dataset. Get creative!

Possible KPIs include (but are not limited to):

• Overall customer satisfaction

• Overall calls answered/abandoned

• Calls by time

• Average speed of answer

• Agents performance quadrant -> average handle time(talk duration) vs calls answered

**• Flow of work**

**Step 1- Upload Data**

The Dataset used for this analysis was presented by PWC\_Switzerland and available at their official website page - [Dataset\_link]

**Step 2-Cleaning data**

Data transformation was done in Power Query and the dataset was loaded into Microsoft Power BI Desktop for modelling. The call canter dataset is given by a table named:

• Call Centre which has 10 columns and 5000 rows of observation

The tabulation below shows the Call centre table with its column names and their description:

| **Column Name** | **Description** |
| --- | --- |
| Call Id | Represents every unique observation in the dataset |
| Agent | Describes the name of the agent |
| Date | Describes the date of the call |
| Time | Represents the time of the call |
| Topic | Describes the topic of the caller |
| Answered | (Y/N) Describes if the call was Answered or not |
| Resolved | Describes if the problem was Resolved or not |
| Speed of answer(in seconds) | Represents the speed of answer in seconds |
| AvgTalkDuration | Represents the average talk duration of call |
| Satisfaction rating | Represents the satisfaction rating of the agent during the call |

**Step 3-Transform data**

Data Cleaning and transformation for the dataset were done in power query as follows:

• Unnecessary columns were removed • Each of the columns in the table was validated to have the correct data type • Unnecessary rows were removed

**Data Visualization**

Data visualization for the datasets was done in Microsoft Power BI Desktop:

• The Call Center Manager Page: Shows KPIs including overall customer satisfaction, overall calls answered/abandoned, calls by time, average speed of answer, agents performance quadrant -> average handle time(talk duration) vs calls answered

**Data Analysis**

Measures used in visualization are:

• %TotalCallsAnswered = DIVIDE([CallsAnswered],COUNT('Call Center Data'[Call Id]))

• CallsAnswered = CALCULATE(COUNT('Call Center Data'[Answered (Y/N)]),'Call Center Data'[Answered (Y/N)]="Y")

• CallsMissed = CALCULATE(COUNT('Call Center Data'[Answered (Y/N)]),'Call Center Data'[Answered (Y/N)]="N")

• IssueResolved = CALCULATE(COUNT('Call Center Data'[Resolved]),'Call Center Data'[Resolved]="Y")

Insights As shown by Data Visualization, It can be deduced that:

• The average satisfaction rating is 3.40

• 4054 calls were answered and 3646 issues resolved

• Jim has the highest satisfaction rating

• The average speed of answer is 67.52 seconds

• Jim has answered total 536 call which is highest whereas Stewart answered lowest number of calls i.e. 477

## ****Task 3-Customer Retention-Customer demographics and insights.****

**•Problem Statement**

The telecom Retention Manager has scheduled a meeting with the engagement partner at PwC to cover these points:

• Customers in the telecom industry are hard-earned: we don’t want to lose them.

• The retention department is here to get customers back in case of termination.

• Currently, we get in touch after they have terminated the contract, but this is reactionary: it would be better to know in advance who is at risk.

• We have done customer analysis with Excel: it has always ended in a dead-end.

• We would like to know more about our customers: visualised clearly so that it’s self-explanatory for our management.

Your colleague, the engagement partner, asks you to do the following tasks:

1. Define proper KPIs
2. Create a dashboard for the retention manager reflecting the KPIs
3. Write a short email to him (the engagement partner) explaining your findings, and include suggestions as to what needs to be changed

Here are some inputs:

• Customers who left within the last month

• Services each customer has signed up for: phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies

• Customer account information: how long as a customer, contract, payment method, paperless billing, monthly charges, total charges and number of tickets opened in the categories administrative and technical

• Demographic info about customers – gender, age range, and if they have partners and dependents

**• Flow of work**

**Step 1- Upload Data**

The Dataset used for this analysis was presented by PWC\_Switzerland and available at their official website page - [Dataset\_link]

**Step 2-Cleaning data**

Data transformation was done in Power Query and the dataset was loaded into Microsoft Power BI Desktop for modelling.The Customer Retention dataset is given by a table named:

• Customer retention which has 23 columns and 7043 rows of observation

The tabulation below shows the Customer retention table with its column names and their description:

| **Column Name** | **Description** |
| --- | --- |
| customerID | Represents the unique number of the customer in the dataset |
| gender | Describes the gender of the customer |
| SeniorCitizen | Describes if the customer is a senior citizen |
| Partner | Describes if the customer has a partner |
| Dependents | Describes if the customer has a dependent |
| tenure | Describes how long as a customer |
| PhoneService | Describes if the customer has registered a phone service |
| MultipleLines | Describes if the customer has registered multiple lines |
| InternetService | Describes if the customer has registered for internet service |
| OnlineSecurity | Describes if the customer has registered for online security |
| OnlineBackup | Describes if the customer has registered for online backup |
| DeviceProtection | Describes if the customer has registered for device protection |
| TechSupport | Describes if the customer has registered for tech support |
| StreamingTV | Describes if the customer has registered to stream tv |
| StreamingMovies | Describes if the customer has registered to stream movies |
| Contract | Describes if the length of the contract of the customer |
| PaperlessBilling | Describes if the customer has registered for paperless billing |
| PaymentMethod | Describes the payment method of the customer |
| MonthlyCharges | Represents the monthly charge incurred by the customer |
| TotalCharges | Represents the total charge incurred by the customer |
| numAdminTickets | Represents the number of admin tickets opened by the customer |
| numTechTickets | Represents the number of tech tickets opened by the customer |
| Churn | Describes if the customer is at risk of churn |

**Step 3-Transform data**

Data Cleaning and transformation for the dataset were done in power query as follows: • Each of the columns in the table was validated to have the correct data type • Unnecessary rows were removed

**Data preparation: -**

**Data Modelling**

After the dataset was cleaned and transformed, it was ready to be modelled. • The official dataset is marked as the 01 CHURN-Dataset table in the dataset.

• Along with this, a separate table- All Measuress is created to store all the measures which we have used in this model.

**Data Visualization**

Data visualization for the datasets was done in Microsoft Power BI Desktop:

• Demographic Info : Shows the male-female distribution,churn by senior citizen,dependents impact,partner impact effect of internet service provider recommendations and a short insight on the report

• Account Info : shows contrct type payment method tenuraity impact on customer count total charges, average monthly charges etc

• Services : shows customer count by streamingTV,customer count by streamingmovies,churn by phoneservice,online security device protection,Tech support etc

• Email / Insights and suggestion :Insights ans suggestions to the engagement partner explaining findings, and include suggestions as to what needs to be changed

**Data Analysis**

Measures used in visualization are:

• Average MonthlyCharges = AVERAGE('01 Churn-Dataset'[MonthlyCharges])

• Average TotalCharges = AVERAGE('01 Churn-Dataset'[TotalCharges])

• churn count = CALCULATE(COUNT('01 Churn-Dataset'[Churn]), ALLSELECTED('01 Churn-Dataset'[Churn]),'01 Churn-Dataset'[Churn] = "yes")

• churn rate % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[Churn]), '01 Churn-Dataset'[Churn] = "yes" ), COUNT('01 Churn-Dataset'[Churn]), 0)

• Dependent in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[Dependents]), '01 Churn-Dataset'[Dependents]="Yes",'01 Churn-Dataset'[Churn]="Yes"), CALCULATE(COUNT('01 Churn-Dataset'[Dependents]),'01 Churn-Dataset'[Churn]="Yes"), 0)

• Device protection in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[DeviceProtection]), '01 Churn-Dataset'[DeviceProtection] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[DeviceProtection]),'01 Churn-Dataset'[Churn]="Yes"),0)

• loyalty = SWITCH(TRUE(),'01 Churn-Dataset'[tenure]<=12,"< 1 year",'01 Churn-Dataset'[tenure]<=24,"< 2 years",'01 Churn-Dataset'[tenure]<=36,"< 3 years",'01 Churn-Dataset'[tenure]<=48,"< 4 years", '01 Churn-Dataset'[tenure]<=60,"< 5 years",'01 Churn-Dataset'[tenure]<=72,"< 6 years")

• Online backup in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[OnlineBackup]), '01 Churn-Dataset'[OnlineBackup] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[OnlineBackup]),'01 Churn-Dataset'[Churn]="Yes"),0)

• Online security in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[OnlineSecurity]), '01 Churn-Dataset'[OnlineSecurity] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[OnlineSecurity]),'01 Churn-Dataset'[Churn]="Yes"),0)

• Partner in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[Partner]),'01 Churn-Dataset'[Partner]="Yes",'01 Churn-Dataset'[Churn]="Yes"), CALCULATE(COUNT('01 Churn-Dataset'[Partner]), '01 Churn-Dataset'[Churn]="Yes"), 0)

• Phone service in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[PhoneService]), '01 Churn-Dataset'[PhoneService] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[PhoneService]),'01 Churn-Dataset'[Churn]="Yes"),0)

• SenioCitizen in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[SeniorCitizenChurn]),'01 Churn-Dataset'[SeniorCitizenChurn]=1,'01 Churn-Dataset'[Churn]="Yes"), CALCULATE(COUNT('01 Churn-Dataset'[SeniorCitizenChurn]),'01 Churn-Dataset'[Churn]="Yes"), 0)

• SeniorCitizen = SWITCH('01 Churn-Dataset'[SeniorCitizenChurn],0,"No",1,"Yes")

• Streaming Movies in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[StreamingMovies]), '01 Churn-Dataset'[StreamingMovies] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[StreamingMovies]),'01 Churn-Dataset'[Churn]="Yes"),0)

• Streaming TV in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[StreamingTV]), '01 Churn-Dataset'[StreamingTV] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[StreamingTV]),'01 Churn-Dataset'[Churn]="Yes"),0)

• Tech Support in % = DIVIDE(CALCULATE(COUNT('01 Churn-Dataset'[TechSupport]), '01 Churn-Dataset'[TechSupport] ="Yes", '01 Churn-Dataset'[Churn]="Yes"),CALCULATE(COUNT('01 Churn-Dataset'[TechSupport]),'01 Churn-Dataset'[Churn]="Yes"),0)

**Insights**

As shown by Data Visualization, It can be deduced that:

• Higher churn rate:

Younger customers (under 3 years).

Customers paying by check or bank transfer.

Customers with paperless billing.

Customers with monthly subscriptions.

• Lower churn rate:

Customers with senior citizen discount.

Customers with multiple lines of service.

Customers with annual subscriptions.

Customers with automatic payments.

• If no Tech Support, Device Protection, and Online Security are provided then the chances of customers churning are high.

• payment method like Electronic check also makes a significant impact on the churning decision.

• There is no relation between gender with churning.

• senior citizens are less likely to churn rate.

• Customers with any dependents or partners are churning less likely, whereas the non-dependents and non-partners customers are more likely to shift.

• Tenure and contract play an important role in determining whether the customer will churn. Customers with monthly contracts i.e. lower tenure will switch frequently.

• Customers with Fiber Optic internet service will churn more compared to DSL internet service holders.

**Suggestions**

• Increasing the basic contract plan from 1 month to 3 months or 6 months can be a good starting point. This will help customers to be with us for a longer tenure.

• Starting special offers or schemes for customers who are single and have no family responsibility. They can become a permanent customer of the company. 'Catch them Young' is key to success here.

• provide some offers to the senior citizens as they are stable customers like overall offer some discounts on long tenure plans.

• Offering basic services like device protection, tech support, and online security should be the primary goal. This will help the customer stay longer with the brand.

­­­­­­­­­­­­­­­­­­­­­

## ­****Task-04 Diversity & Inclusion****

**•Problem Statement**

Task is to do the following:

• Define relevant KPIs in hiring, promotion, performance and turnover, and create a visualisation. Write what you think some root causes of their slow progress might be.

• Currently, we get in touch after they have terminated the contract, but this is reactionary: it would be better to know in advance who is at risk.

Calculating the following measures could help to define proper KPIs:

• # of men

• # of women

•# of leavers

• % employees promoted (FY21)

• % of women promoted

• % of hires men

• % of hires women

• % turnover

• Average performance rating: men

• Average Performance rating: women

**• Flow of work**

**Step 1- Upload Data**

The Dataset used for this analysis was presented by PWC\_Switzerland and available at their official website page - [Dataset\_link]

**Step 2-Cleaning data**

Data transformation was done in Power Query and the dataset was loaded into Microsoft Power BI Desktop for modelling.The Customer Retention dataset is given by a table named:

• Cdiversity and inclusion which has 32 columns and 500 rows of observation

The tabulation below shows the Customer retention table with its column names and their description:

| **Column Name** | **Description** |
| --- | --- |
| Employee ID | Represents the unique number of the employee in the dataset |
| Gender | Describes the gender of the employee |
| Job Level after FY20 promotions | Describes the job level of the employee after being promoted in FY20 |
| New hire FY20? | Describes if the employee is a new hire in FY20 |
| FY20 Performance Rating | Represents the performance rating of the employee in FY20 |
| Promotion in FY21? | Describes if the employee is being promoted in FY21 |
| In base group for Promotion FY21 | Describes if the employee is being selected for promoted in FY21 |
| Target hire balance | Describes the target hire balance of the employee |
| FY20 leaver? | Describes if the employee is a leaver in FY20 |
| In base group for turnover FY20 | Describes if the employee is in a group for turnover in FY20 |
| Department @01.07.2020 | Describes the department each employee belongs to as at January 7, 2020 |
| Leaver FY | Describes if the employee is a leaver in a FY |
| Job Level after FY21 promotions | Describes the job level of the employee after being promoted in FY21 |
| Last Department in FY20 | Describes the last department each employee belongs in FY20 |
| FTE group | Describes if the employee belongs to a FTE group |
| Time type | Describes the contract type employee |
| Department & JL group PRA status | Describes the department and JL group PRA status of the employee |
| Department & JL group for PRA | Describes the department and JL group PRA of the employee |
| Job Level group PRA status | Describes the job level group PRA status of the employee |
| Job Level group for PRA | Describes the job level group PRA of the employee |
| Time in Job Level @01.07.2020 | Describes the time in job level of the employee |
| Job Level before FY20 promotions | Describes the job level employee before being promoted in FY20 |
| Promotion in FY20? | Describes if the employee is being promoted in FY20 |
| FY19 Performance Rating | Describes the performance rating of the employee in FY19 |
| Age group | Describes the age group of the employee |
| Age @01.07.2020 | Represents the age of the employee as at January 07, 2020 |
| Nationality 1 | Describes the nationality of the employee in state level |
| Region group: nationality 1 | Describes the nationality of the employee in country level |
| Broad region group: nationality 1 | Describes the nationality of the employee in regional level |
| Last hire date | Describes the last hire date of the employee |
| Years since last hire | Represents the number of years since last hire of the employee |
| Rand | generates random number for each entry in the dataset |

**Step 3-Transform data**

Data Cleaning and transformation for the dataset were done in power query as follows: • Each of the columns in the table was validated to have the correct data type. • Unnecessary rows were removed. • Unnecessary rows were filtered.

**Data preparation: -**

**Data Modelling**

• After the dataset was cleaned and transformed, it was ready to be modelled.

• Along with this, a separate table- Basic Measuress is created to store all the measures which we have used in this model.

MEN\_new\_HIRES - it's dim table which is created to store all the measures related to men category in this model.

women\_new hires - it's dim table which is created to store all the measures related to women category in this model

**Data Visualization**

Data visualization for the datasets was done in Microsoft Power BI Desktop:

• By Department: This section show the composition of the workforce by department, highlighting the diversity of employees across different departments. • By Age Group: This section display the age distribution of the employees, potentially indicating any potential age-related biases in hiring or promotion practices. • By Region: This section show the geographical distribution of the workforce, indicating whether the company has a diverse workforce across different regions. • By Gender: This section show the gender composition of the workforce, highlighting the representation of different genders within the company. • Department: This section shows the breakdown of employees across different departments. • Age Group: This section displays the age distribution of the employees. • Region Group: This section shows the geographical distribution of the workforce.

**Data Analysis**

Measures used in visualization are:

• # after hire = 'Pharma Group AG'[# men]+[# women]

• # before hire = CALCULATE(COUNT('Pharma Group AG'[Employee ID]),'Pharma Group AG'[In base group for turnover FY20]="Y")

• # Leavers = CALCULATE(COUNT('Pharma Group AG'[FY20 leaver?]), 'Pharma Group AG'[FY20 leaver?]="Yes")

• # Men = CALCULATE(COUNT('Pharma Group AG'[Gender]), 'Pharma Group AG'[Gender]="Male")

• # Women = CALCULATE(COUNT('Pharma Group AG'[Gender]), 'Pharma Group AG'[Gender]="Female")

• % employees promoted = DIVIDE(CALCULATE(COUNT('Pharma Group AG'[In base group for Promotion FY21]),'Pharma Group AG'[Promotion in FY21?]="Yes"), CALCULATE(COUNT('Pharma Group AG'[Promotion in FY21?]), 'Pharma Group AG'[In base group for Promotion FY21]="Yes"))

• % hires men = DIVIDE('Pharma Group AG'[# men], 'Pharma Group AG'[# men]+[# women])

• % hires women = DIVIDE('Pharma Group AG'[# women], 'Pharma Group AG'[# men]+[# women])

• % turnover = [# leavers] / 'Pharma Group AG'[average # employee]

• % women promoted = DIVIDE(CALCULATE(COUNT('Pharma Group AG'[In base group for Promotion FY21]),'Pharma Group AG'[Promotion in FY21?]="Yes", 'Pharma Group AG'[Gender]="Female"), CALCULATE(COUNT('Pharma Group AG'[Promotion in FY21?]),'Pharma Group AG'[Promotion in FY21?]="Yes"))

• average # employee = ([# before hire] + 'Pharma Group AG'[# after hire])/2

• Avg performamnce (men) = CALCULATE(AVERAGE('Pharma Group AG'[FY20 Performance Rating]),'Pharma Group AG'[Gender]="Male")

• Avg performamnce (women) = CALCULATE(AVERAGE('Pharma Group AG'[FY20 Performance Rating]),'Pharma Group AG'[Gender]="Female")

**Insights**

####As shown by Data Visualization, It can be deduced that:

• Overall, the company seems to have a good gender balance in terms of hiring, with women making up just under half (41%) of new hires. However, there is a significant disparity in promotion rates, with only 25% of promotions going to women. This suggests that there may be barriers in place that are preventing women from being promoted at the same rate as men.

• The turnover rate is higher for women than for men. This could be due to a number of factors, including the gender pay gap, a lack of opportunities for advancement, or a hostile work environment.

• The average performance rating of the employees decreased from to in FY20..

• Maximum hiring of employees is done from Switzerland ,France & Germany respectively, hence in order to increase diversity need to hire talented employees from different part of globe.

• The slow progress in the executive level is of the fact that only less than 20% female is promoted to this managerial and executive roles.

• Employee promotion rate is increase by 3% in FY21 than FY20.

• FY20 Hires vs FY21 Job Level: While the number of females hired increased slightly from FY20 (40) to FY21 (41), the number of females in senior positions (director and above) remained the same (3) in both years. This indicates a potential gap in promoting women to higher-level positions.

• Women are more likely to be in lower-level positions, such as junior officers and managers, while men are more likely to be in higher-level positions, such as directors and executives. This suggests that there may be a lack of opportunities for women to progress to senior leadership positions.

**Suggestions**

####To ensure progress in diversity and inclusion in the executive level;

• More women should be hired and most especially promoted because the gap in the ratio of men to women is quite large.

• Forthe Executive and Director position ,female employee count as well as the promotion count is too low compared to male employee hence more women should be hired as well as promoted.

• Age group 30-39 has more rate of promotion compared to 40-49 age group, experience should be consider as one of the the one of the criteria for promotion checklist.