Headcount Data Analysis by John Burnett

Abstract

The HR Department wanted to create a new employee database with the view to analyze the data on a daily, weekly and monthly basis.

Introduction

The main areas of concern are the attrition rates and if there are any departments or work bases that stand out.

To build a robust analysis model using the cross-industry process for data mining (CRISP-PM) [1], we will look at the following phases:

1. Understanding of context: Introduction
2. Data understanding: Initial Exploration
3. Data preparation: Pre-processing
4. Modeling: Selection of the model
5. Evaluation: Conclusion
6. Deployment

Object and scope:

For this project a employee database will be built using MYSQL along and PowerBI will be used to build the visualizations to answer the main questions raised by HR:

1. What is the current headcount within the company.
2. What is the headcount, joiners and leavers on a monthly basis.
3. What is the current level of attrition?
4. Are there and areas of concern such as high attrition rates in a employee base, a particular department or any other factors?

Data source:

The main data source will be the newly created MYSQL database with an addition of a date dimension table (excel) and in order to include a visual map to represent the bases, a list of airport codes with their respective longitude and latitude on a csv file will be used.

The data used to create the database in this scenario is random data.

Input attributes:

The MYSQL data set contains 1000 examples with 16 attributes as follows:

Employee Number
First Name
Surname
Gender
Start date
Base
Department
Job Title
Rank
Contract Type
Fleet
FTE
Exit Date
Method of Leaving
New Employer
Salary

Analysis

Initial Exploration:

After building the database, the data was collected in PowerBI Desktop to see how the data looked, are there errors or erroneous data within and if a clean up of the data structure was required?

Two additional data sources were added:

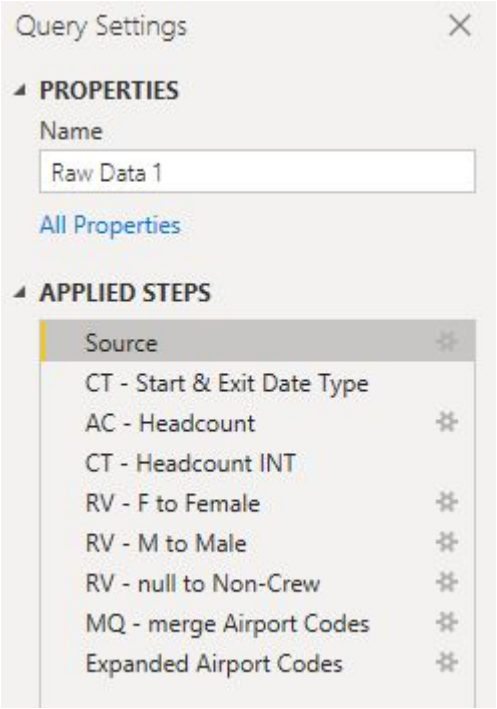
1. Date - excel file to use as a date dimension
2. Airport codes - csv file to add longitude and latitude used in maps

Additional applied steps were added in order to make the data more usable:

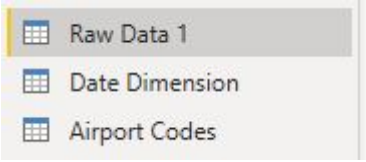
Start and Exit dates change to date type
Additional column Headcount was added to use in summing and data type integer
Replacing of F and M to Female and Male
Replacing of null to Non-Crew
Merge of Airport Codes file and then expanding the fields of Longitude and Latitude

Most of the applied steps have been renamed for future use if amendments are required.

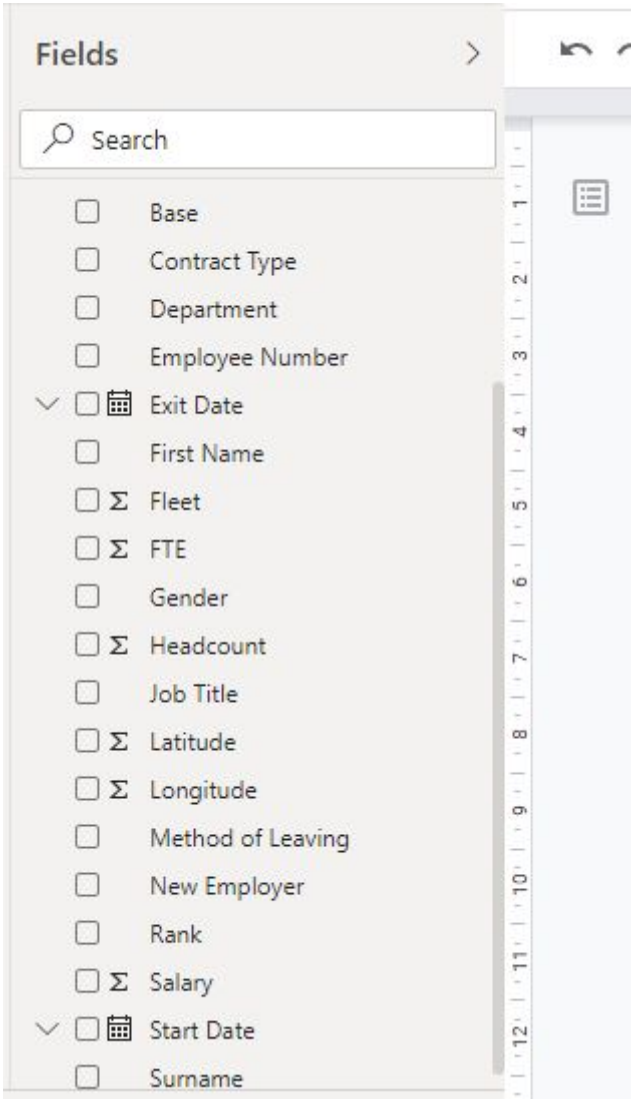
After the above there are now three queries that we can start to use to model our data.



After the above there are now three queries that we can start to use to model our data.



This will give us the following fields:



A JSON file is added to create a report theme that contains the company corporate colours, text sizes and colours plus a background image.

To test that the data being used is correct a table of data is created to see if the total count of employees is correct plus the date formats are correct and at an over-view the information looks fine.

Employee Number	First Name	Surname	Base	Start Date	Department	Contract Type	Gender	Job Title	Rank	Fleet	FTE	Exit Date	Method of Leaving	New Employer	Headcount
138130	Lauren	Mcgee	BFS	10/07/1987	OPS	Direct	Female	Cabin	JCC		1				1
149763	Dudley	Melton	KIR	10/07/1987	Engineering	Direct	Male		Non-Crew		1				1
287114	Joanna	Perkins	KIR	10/07/1987	OPS	Direct	Female	Cabin	JCC		1				1
288820	Eric	Mccullough	ORK	10/07/1987	OCC	Direct	Male		Non-Crew		1				1
349751	Shelton	Harrison	DUB	10/07/1987	Engineering	Direct	Male		Non-Crew		1				1
381597	Edwardo	Stone	BFS	10/07/1987	OPS	Direct	Male	Pilot	FO	737	1				1
395792	Miguel	Finley	BFS	10/07/1987	OCC	Direct	Male		Non-Crew		1	30/07/2020	Resigned	AA	1
422323	Wilma	Gray	KIR	10/07/1987	Commercial	Direct	Female		Non-Crew		1				1
431018	Rickie	Alston	DUB	10/07/1987	OCC	Direct	Male		Non-Crew		1				1
434723	Melanie	Zimmerman	ORK	10/07/1987	Commercial	Direct	Female		Non-Crew		1				1
498361	Elise	Zimmerman	DUB	10/07/1987	IT	Direct	Female		Non-Crew		1				1
505758	Garland	Watkins	ORK	10/07/1987	HR	Direct	Male		Non-Crew		1				1
520092	Lamont	Woods	KIR	10/07/1987	OPS	Agency	Male	Cabin	CS		1				1
535812	Lavonne	Carney	DUB	10/07/1987	OPS	Direct	Female	Cabin	JCC		1				1
539548	Yong	Poole	KIR	10/07/1987	OCC	Direct	Male		Non-Crew		1				1
551251	Frederick	Graham	KIR	10/07/1987	HR	Direct	Male		Non-Crew		1				1
572348	Selma	Floyd	ORK	10/07/1987	OPS	Direct	Female	Pilot	CP	737	1	15/02/2019	Resigned	FR	1
616656	Isiah	Hubbard	DUB	10/07/1987	HR	Direct	Male		Non-Crew		1				1
621988	Callie	Snider	ORK	10/07/1987	IT	Direct	Female		Non-Crew		1				1
648744	Edmund	Blankenship	BFS	10/07/1987	Engineering	Direct	Male		Non-Crew		1				1
673049	Cole	Jensen	BFS	10/07/1987	Commercial	Direct	Male		Non-Crew		1	05/12/2019	Resigned	EI	1
747647	Trisha	Mueller	DUB	10/07/1987	OPS	Direct	Female	Pilot	FO	737	1				1
809726	Bryce	Cox	KIR	10/07/1987	IT	Direct	Male		Non-Crew		1				1
Total											**994**				**1000**

Creation of Measures:

Measures are created to increase the number of fields required for the report to show the analysis that was initial asked for, here are a few examples:

To get a monthly headcount that reflects the start and exit dates of employes:

```
101.Monthly Headcount =
VAR _FirstDayCurrentMonth = MIN ( 'Date Dimension'[Date] )
VAR _LastDayCurrentMonth = MAX ( 'Date Dimension'[Date] )
RETURN
    CALCULATE (SUM ( 'Raw Data 1'[Headcount]),
        FILTER ('Raw Data 1','Raw Data 1'[Start Date] <= _LastDayCurrentMonth
            && ('Raw Data 1'[Exit Date] > _LastDayCurrentMonth
                || ISBLANK ( 'Raw Data 1'[Exit Date] ))))
```

To show monthly joiners:

```
102.Monthly Joiners =
VAR _FirstDayCurrentMonth =
    MIN ( 'Date Dimension'[Date] )
VAR _LastDayCurrentMonth =
    MAX ( 'Date Dimension'[Date] )
RETURN
    CALCULATE (SUM ( 'Raw Data 1'[Headcount]),
        FILTER ('Raw Data 1', 'Raw Data 1'[Start Date] <= _LastDayCurrentMonth
            && ('Raw Data 1'[Start Date] >= _FirstDayCurrentMonth ))))
```

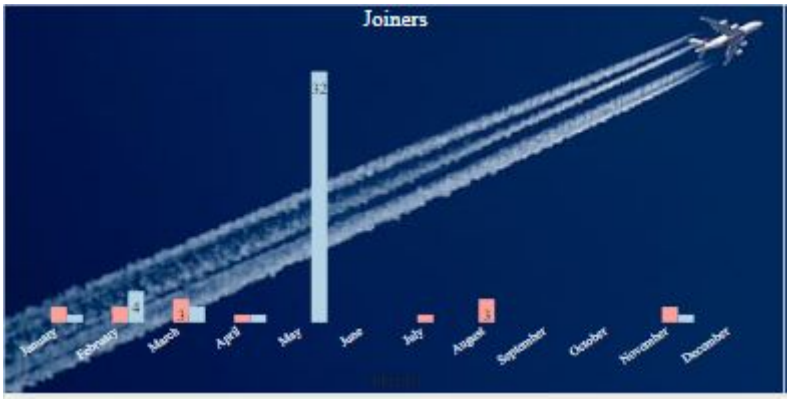
To show monthly leavers:

```
103.Monthly Leavers =
VAR _FirstDayCurrentMonth = MIN ( 'Date Dimension'[Date] )
VAR _LastDayCurrentMonth =  MAX ( 'Date Dimension'[Date] )
RETURN
    CALCULATE (SUM ( 'Raw Data 1'[Headcount] ),
        FILTER ('Raw Data 1', 'Raw Data 1'[Exit Date] <= _LastDayCurrentMonth
                    && ('Raw Data 1'[Exit Date] >= _FirstDayCurrentMonth)))
```

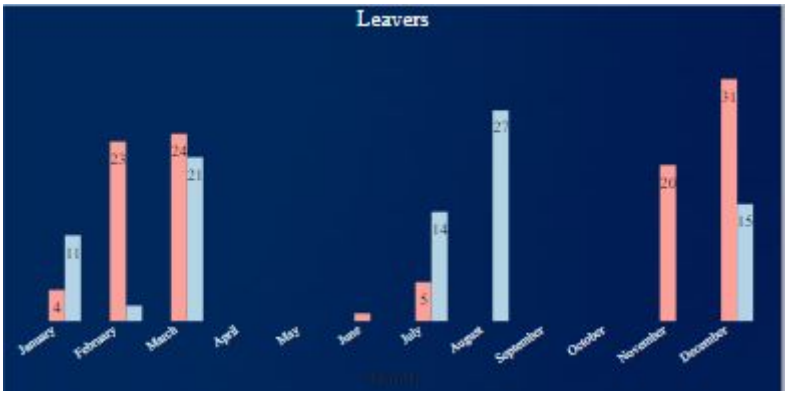
With such measures and the fields, the report can be created, the first page will be visuals (clustered column charts) to show the last 3 years at a monthly level of the total headcount, the number of leavers and joiners.

Headcount (monthly level over 3 years)

Joiners (monthly level over 3 years)



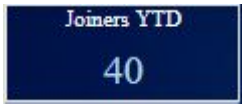
Leavers (monthly over 3 years)



To aid in the quick view of certain numbers or elements, cards are added:

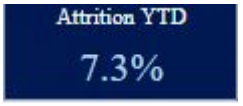
The employee headcount at 01/01/2020 (when the year changes it will automatically go to 01/01/2021

The current headcount year to date



The joiners year to date

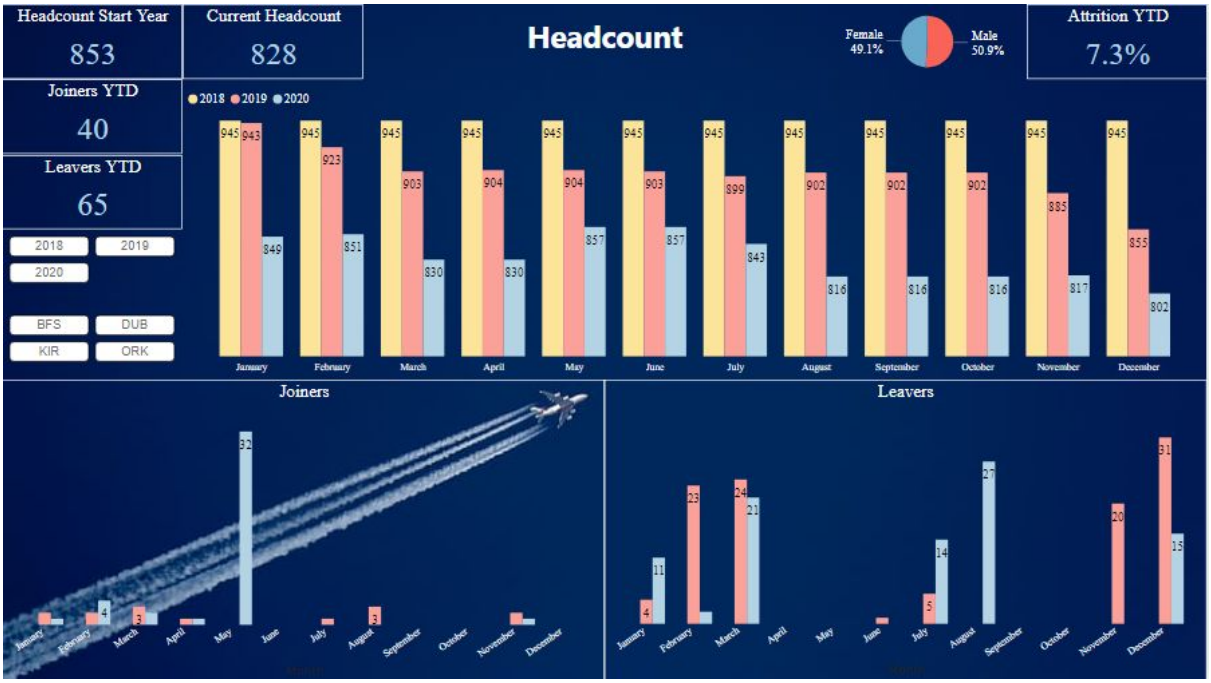
The leavers year to date



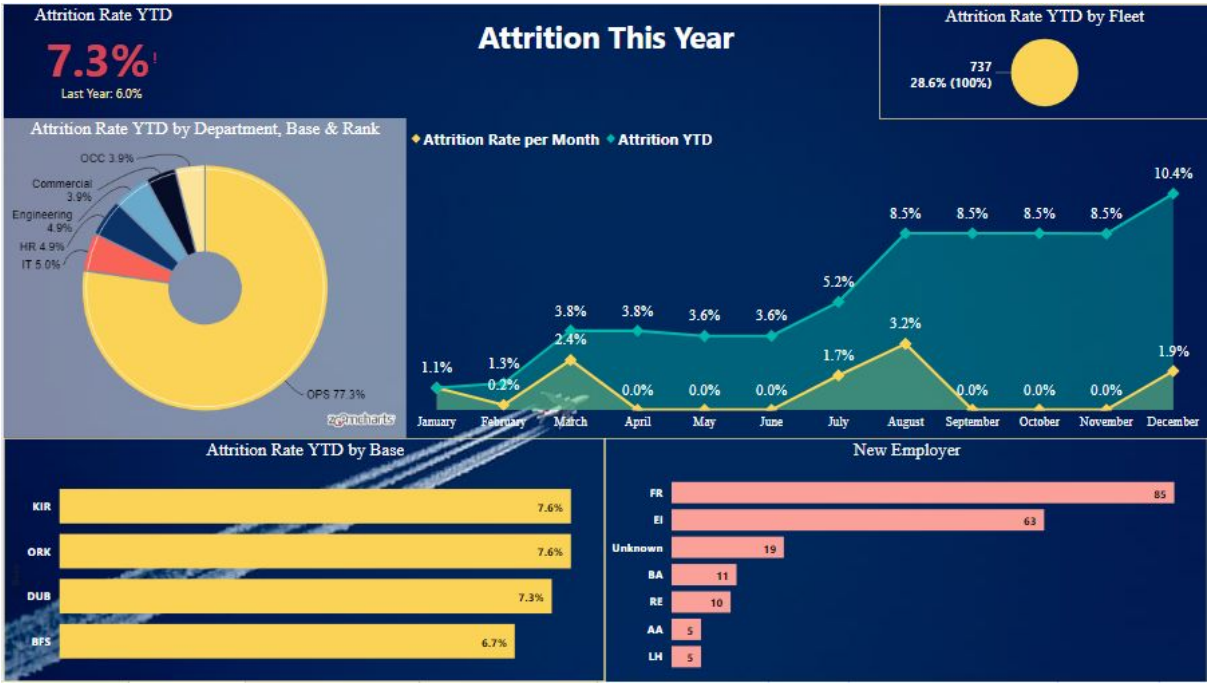
The level of attrition year to date decided by HR as “Leavers divided by (headcount at start of year plus joiners year to date)”

To show the gender mix a Pie chart is added.

The Headcount page is complete and looks like this



To show the breakdown of attrition levels a second page is developed with the following



At a glance we can see the problem areas, the months that have increased attrition levels compared to others, the departments most affected by this and what other companies staff are moving on to.

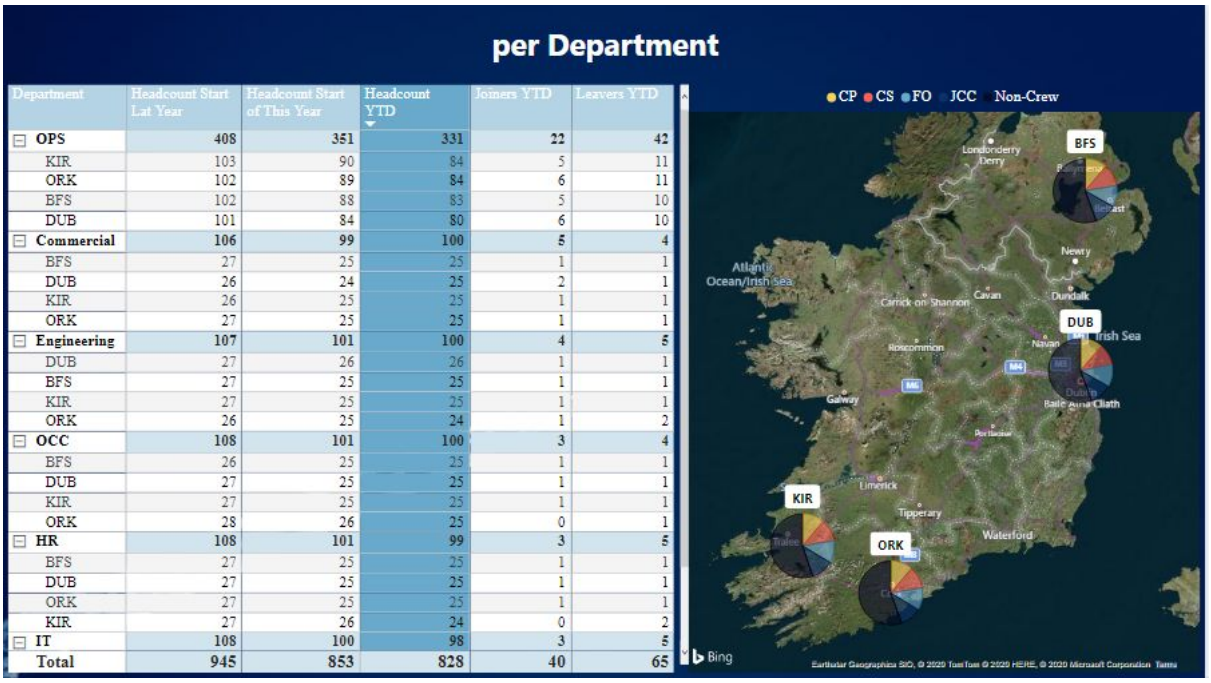
Additional pages (Joiners This Year and Leavers This Year) allow the end user to check how many joiners and leavers there are in each base and department.

Base	January	February	March	April	May	June	July	August	September	October	November	December
⊞ ORK		1			8							
⊞ KIR			1	1	7							
⊞ DUB		2	1		9						1	
⊞ BFS	1	1			8							
Total	**1**	**4**	**2**	**1**	**32**						**1**	

By drilling down it expands further.

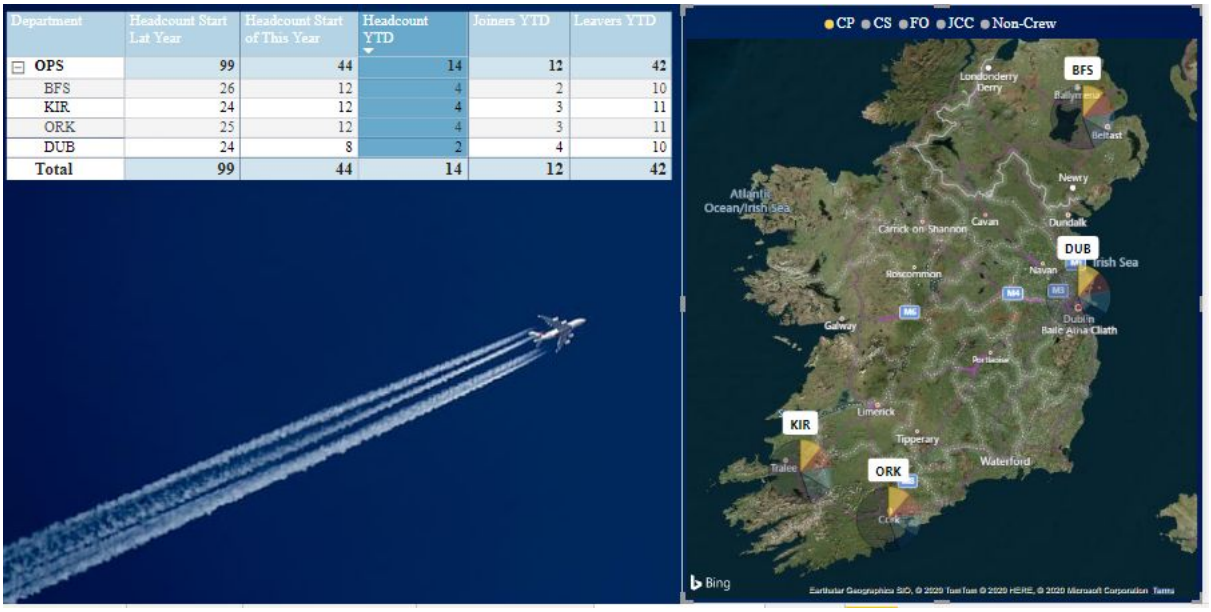
Base	January	February	March	April	May	June	July	August	September	October	November	December
⊟ **ORK**		**1**			**8**							
OPS		1			5							
OCC												
IT												
HR					1							
Engineering					1							
Commercial					1							
⊟ **KIR**			**1**	**1**	**7**							
OPS			1	1	3							
OCC					1							
IT					1							
HR												
Engineering					1							
Commercial					1							
⊟ **DUB**		**2**	**1**		**9**						**1**	
OPS		2	1		3						1	
OCC					1							
IT					1							
HR					1							
Engineering					1							
Commercial					2							
⊟ **BFS**	**1**	**1**			**8**							
OPS	1	1			3							
OCC					1							
IT					1							
HR					1							
Engineering					1							
Total	**1**	**4**	**2**	**1**	**32**						**1**	

The per Department page shows at a glance each department, the bases for each and the previous employment levels along side the current one with an interactive map which can be broken down to the rank of each employee.



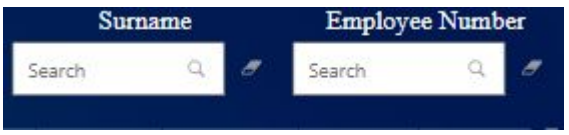
Department	Headcount Start Lat Year	Headcount Start of This Year	Headcount YTD	Joiners YTD	Leavers YTD
⊟ OPS	408	351	331	22	42
KIR	103	90	84	5	11
ORK	102	89	84	6	11
BFS	102	88	83	5	10
DUB	101	84	80	6	10
⊟ Commercial	106	99	100	5	4
BFS	27	25	25	1	1
DUB	26	24	25	2	1
KIR	26	25	25	1	1
ORK	27	25	25	1	1
⊟ Engineering	107	101	100	4	5
DUB	27	26	26	1	1
BFS	27	25	25	1	1
KIR	27	25	25	1	1
ORK	26	25	24	1	2
⊟ OCC	108	101	100	3	4
BFS	26	25	25	1	1
DUB	27	25	25	1	1
KIR	27	25	25	1	1
ORK	28	26	25	0	1
⊟ HR	108	101	99	3	5
BFS	27	25	25	1	1
DUB	27	25	25	1	1
ORK	27	25	25	1	1
KIR	27	26	24	0	2
⊟ IT	108	100	98	3	5
Total	945	853	828	40	65

If “CP” is clicked in the map legend, both the map and the chart will show only the levels of CP’s.

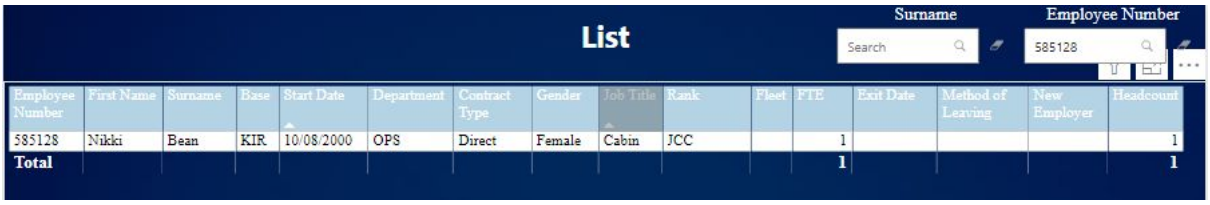


Department	Headcount Start Lat Year	Headcount Start of This Year	Headcount YTD	Joiners YTD	Leavers YTD
⊟ OPS	99	44	14	12	42
BFS	26	12	4	2	10
KIR	24	12	4	3	11
ORK	25	12	4	3	11
DUB	24	8	2	4	10
Total	99	44	14	12	42

For the list of all past, current and new employees, the last page has new search bars added.



For example



Employee Number	First Name	Surname	Base	Start Date	Department	Contract Type	Gender	Job Title	Rank	Fleet	FTE	Exit Date	Method of Leaving	New Employer	Headcount
585128	Nikki	Bean	KIR	10/08/2000	OPS	Direct	Female	Cabin	JCC		1				1
Total											**1**				**1**

Conclusion

Results:

After completing the PowerBI report with the data input from various sources and constructing several measures and columns, we have an overview of current staffing levels across the company and where there might be issues.

We have answered the questions raised by HR that requested the report to be built.

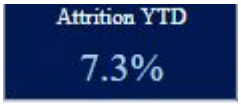
1. What is the current headcount within the company

The current headcount (17/08/2020) is 828 and when the report is opened on another day/date, the dynamics of the report will accurately reflect the current date and headcount.

2. What is the headcount, joiners and leavers on a monthly basis.

On the Headcount page the above is shown

3. What is the current level of attrition?



The level at present is 7.3% and will show the correct % whenever the report is reviewed at a later date.

4. Are there any areas of concern such as high attrition rates in an employee base, a particular department or any other factors?

From the Attrition page we can see that the following areas should be of concern:

a. Employees working in the OPS department currently have the highest level of attrition.

b. The Base of KIR (or Kerry) and ORK (Cork) have the greatest level of attrition.
c. The company ais losing staff mainly to the airline FR.
d. We have only lost pilots rated on the 737 due to attrition.
e. Compared to the same period last year, the attrition rate is higher (7.3% compared to 6.0%).

References

[1] Smart Visión Group. What is the CRISP-DM Methodology. [https://www.sv-europe.com/crisp-dm-methodology/]