QuantSpark

Case Study

QuantSpark Second Round Analyst Assessment

Information 01

Information 02

Information 03

Contents

Congratulations on Advancing to the Case Study!

- This round will consist of a preparing a presentation, based on analysis of a sample dataset that you should have received.

- This presentation should be written as if you were presenting the analysis and findings back to the client.

- The task is all about eliciting insight and developing a compelling narrative driven by strong argument and graphical evidence.

Put yourself in the shoes of a QuantSpark Analyst: You will help answer a client's question using data science

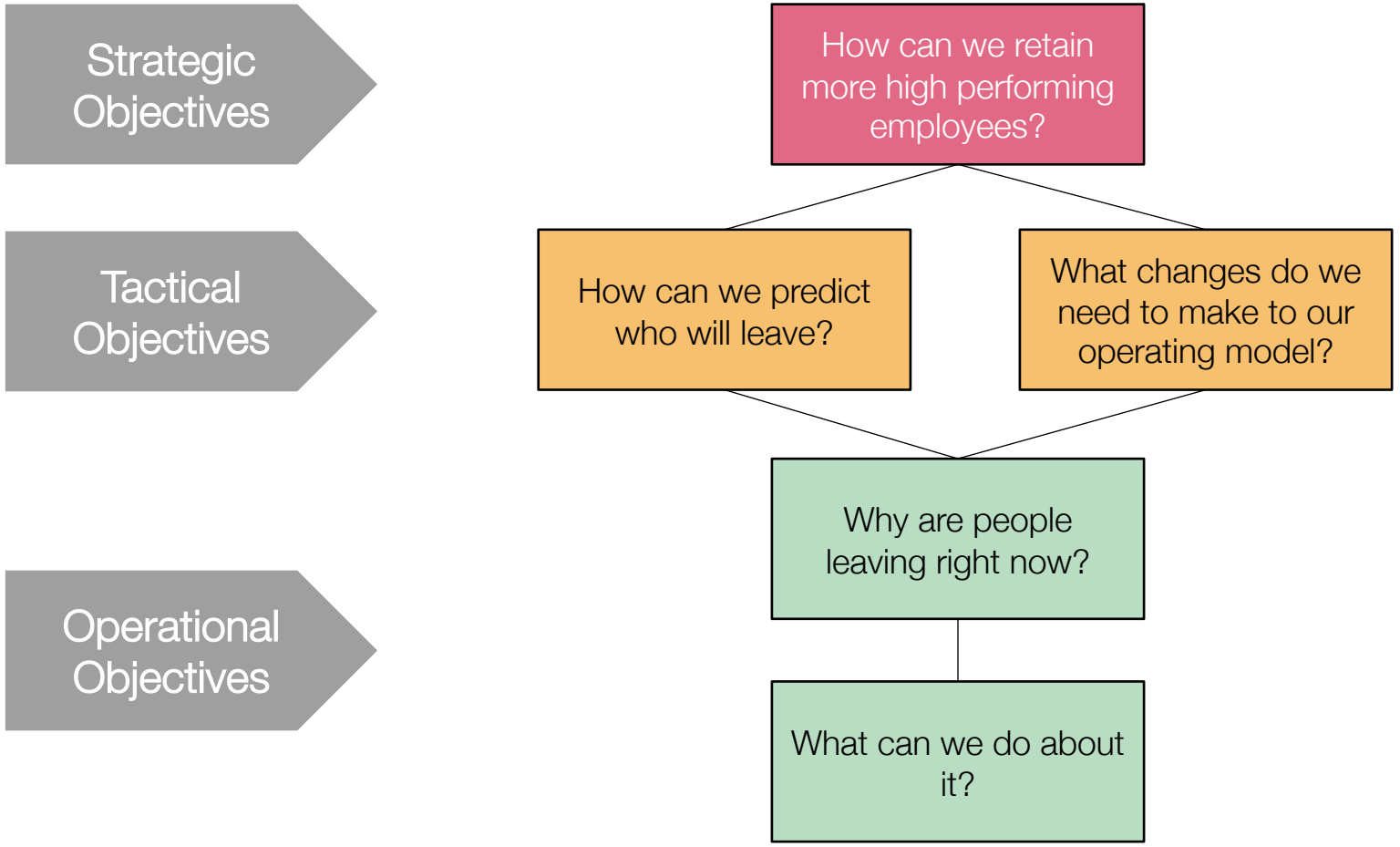
Client Brief

- The client is concerned about retention of its high performing employees and has asked you to investigate how it can improve retention of such colleagues
- The client is also interested in being able to predict exactly which of its employees is most at risk of leaving
- To aid in this task, the client has provided some data which includes attributes of its current and past employees

This exercise is an opportunity for you to showcase your data science, technical and commercial expertise to us

Contents

To help you answer the brief, you have been provided with a HR dataset that contains employee attributes and flags past leavers



The dataset has 20 “dimensions” and 1 outcome; there are 1,500 rows so the file can be examined in Excel

PerformanceRating	TotalWorkingYears	YearsAtCompany	YearsSinceLastPromotion	Left
1	1	1	0	No
2	8	7	3	No
3	7	0	0	No
3	23	8	0	No
1	7	7	0	No
1	1	0	0	Yes
5	13	11	4	No
5	11	10	1	No
5	6	5	0	No
2	12	7	0	No
2	28	10	1	No
3	13	12	5	No
2	12	9	4	No
1	10	8	7	No
3	4	2	2	No
3	34	33	15	No
2	4	3	1	No
3	7	6	1	No
5	3	0	0	Yes
3	29	8	0	No
2	5	5	1	No
4	9	5	1	No
3	18	4	0	Yes
2	17	0	0	No
1	8	6	1	Yes
5	3	3	1	No
2	6	5	0	No
4	5	3	0	No
4	19	2	2	No
1	5	3	1	No
4	8	1	0	No
2	8	5	0	No

There are 20 variables - “Left” is the outcome we’re analysing

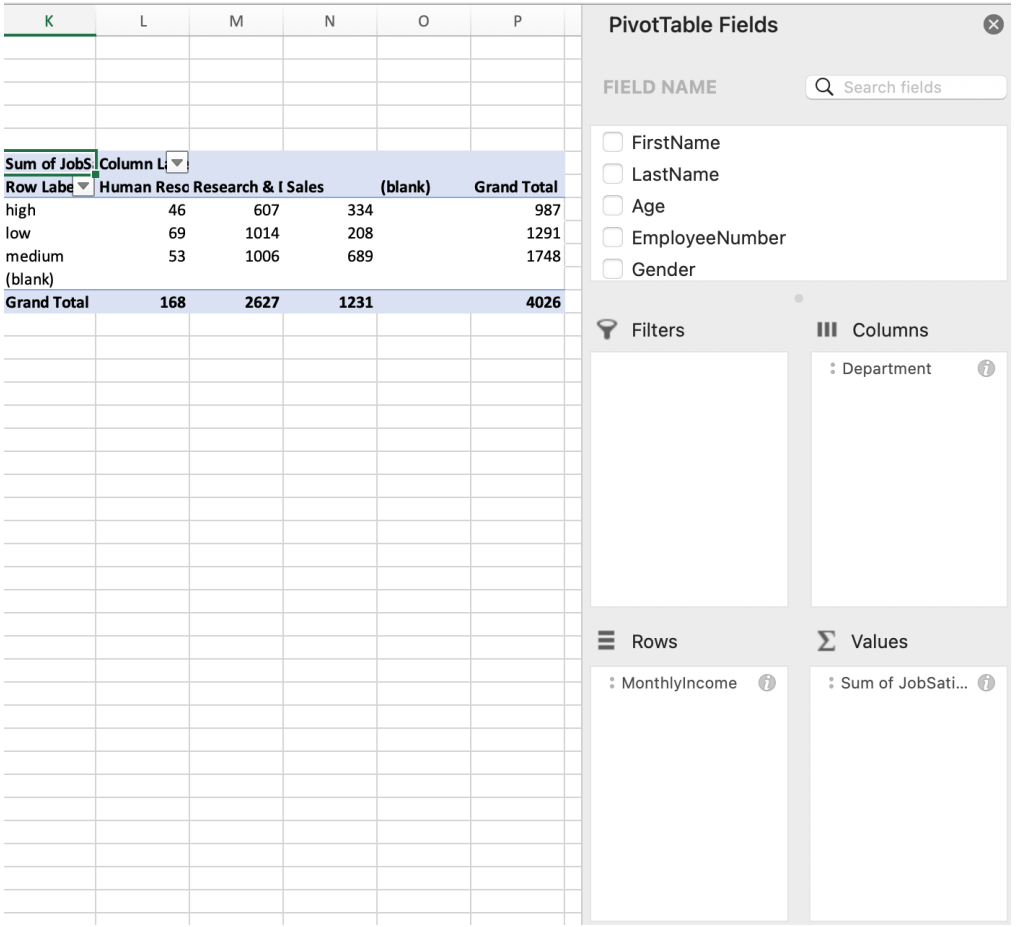
Age	The age of employees
Gender	The gender of employees
Monthly Income	Monthly Income, either low, medium or high
Department	The department the employee works in
NumCompanies Worked	The number of companies the employee worked at before current
Over 18	Whether the employee is over 18 or not
Workingfrom home	Whether or not the employee has the option to work from home
BusinessTravel	How often the employee travels for business: Rarely, Frequently or never
DistanceFrom Home	The distance the employee lives from the office
StandardHours	The number of hours each employee works in a standard week

There are 20 variables - “Left” is the outcome we’re analysing

JobSatisfaction	The satisfaction of the employee from their last evaluation, ranging from 1-4
ComplaintFiled	Whether or not the employee has filed a complaint
Complaint Resolved	Whether or not the employee’s complaint has been resolved
Complaint Years	The number of years since the complaint was filed
PercentSalary Hike	The last percentage salary increase the employee received
Performance Rating	The last performance rating the employee received after evaluation, 1-5
TotalWorking Years	Total number of years the employee has worked
YearsAt Company	Number of years employee has spent at company
YearsSince Promotion	Number of years since employee has been promoted
Left	Whether or not the employee has left the company or not

Building ad hoc charts and Pivot Tables in Excel is a quick way of understanding data and developing hypotheses initially

PerformanceRating	TotalWorkingYears	YearsAtCompany	YearsSinceLastPromotion	Left
1	1	1	0	No
2	8	7	3	No
3	7	0	0	No
3	23	8	0	No
1	7	7	0	No
1	1	0	0	Yes
5	13	11	4	No
5	11	10	1	No
5	6	5	0	No
2	12	7	0	No
2	28	10	1	No
3	13	12	5	No
2	12	9	4	No
1	10	8	7	No
3	4	2	2	No
3	34	33	15	No
2	4	3	1	No
3	7	6	1	No
5	3	0	0	Yes
3	29	8	0	No
2	5	5	1	No
4	9	5	1	No
3	18	4	0	Yes
2	17	0	0	No
1	8	6	1	Yes
5	3	3	1	No
2	6	5	0	No
4	5	3	0	No
4	19	2	2	No
1	5	3	1	No
4	8	1	0	No
2	8	5	0	No



Contents

After analysing the data, we would like you to distil the key trends and insights into a PowerPoint presentation

Get the dataset

- Should be in the email this slide deck was in!

1-3 hours of analysis

- Identify trends
- Try to dig deeper if you can
  - Segment groups
  - Develop a predictive model to predict leavers
- Visualise key insights

Present findings

- Create a short PowerPoint presentation(10-15 slides)
- Make sure it's message-driven: Focus on answering the **"so-what"** questions

If you want to learn more, there's lots of (free) material online

- Introduction to Data Analysis using Excel
  https://www.edx.org/course/introduction-data-analysis-using-excel-microsoft-dat205x-0#
- Making Smarter, More Persuasive Data Visualizations
  http://spotfire.tibco.com/assets/blt25ae9ecd4ca3da2e/hbr-20160330-dataviz.pdf
- Analyzing and Visualizing Data with Excel
  https://www.edx.org/course/analyzing-visualizing-data-excel-microsoft-dat206x-4#!
- Statistical Thinking for Data Science and Analytics
  https://www.edx.org/course/statistical-thinking-data-science-columbiax-ds101x-0
- Introduction to Probability - The Science of Uncertainty
  https://www.edx.org/course/introduction-probability-science-mitx-6-041x-2
- University of California Berkeley's Visualisation (course slides)
  http://vis.berkeley.edu/courses/cs294-10-sp11/wiki/index.php/CS294-10_Visualization
- Harvard's Statistics 110: Probability (YouTube lectures)
  http://projects.iq.harvard.edu/stat110/youtube
- Tableau training videos
  https://www.tableau.com/learn