ADS-AInterviews and Notes

korte regel

Group B  
11 september 2020

# **Table of Contents**

[**Table of Contents**](#_a25bt2p57vp5) **1**

[**1.Interview: Q&A - 9/11/2020**](#_g1fuvd91czrh) **2**

[1.1.Notes](#_nyon7tc6j8be) 2

[1.2.Class Questions](#_g5rooeosuoe2) 2

[**2.Conversation: Danny & Simona - 9/17/2020**](#_9nrchlqwtg44) **7**

[2.1.Notes](#_w7o95tmx3a0f) 7

# 1.Interview: Q&A - 9/11/2020

## **1.1.Notes**

Rob Prop, Informa is an IT consultancy company that is here for more than 30 years, Rob has been working here for 20 years. The company does data related challenges and started as a data company and has been expanding in things like data science and data security. There is Informa Belgium and Informa Netherlands. The Informa in Belgium does not have technological employees working there. So the Dutch department does their technological work.

Informa also does conceptual work for customers. Informa works with programs like Jupyter Notebook and R.

The company agreed with special conditions but didn't have time to participate, they wanted informa to be in contact with Fontys. The people that work with the data should work with as much respect as possible with the data because they want to stay anonymized, but are very interested in the results.

During the creation of the dataset they removed some parts of the data :)

What makes it easy is that the number of columns is pretty low in the dataset. Everything in the dataset is made in the same format. The information comes from the customers. The customers get this data from their customers.

The contact has a bit of knowledge of the dataset.

## **1.2.Class Questions**

**The customer can pay part of the price. There is a column called contribution in the dataset?**

This is what the contact has mentioned. The contribution is what the customer pays themselves and is not included in the price. The price column is the total price and the contribution is what the customer pays themselves. The contribution is different per customer based on the insurance company involved. Some insurance companies have their maximum contribution that has to be paid. So it could change depending on the month how much has to be paid.

**Which department has the most use of analytics?**

A couple of years ago it was marketing, but now it is more used in production and logistics. Customers want more and more a 360 degree view of their customers. They want to see the customer as a customer for the whole company.

**Do you use insight or guides for future strategies?**

Yes, they do that more and more. For one of their bank customers they use mortgages. So which mortgages are risky and which are not risky. So where did they do it, good or not good. The most important thing is to infuse the model in the operational system. So that they could use the same model for an extra fee.

**Regarding the pharmacies, how do you use it there?**

They don't use data science yet for this customer, but want to do things like looking at the totals of the pharmacies. They are dealing with insurance companies and distributors to negotiate the prices. So predict things like how many products do they need and predict time. Another could be benchmarking or looking at what kind of products have an atc code or dont have one. The ones with the code a pre subscribes medications. Like a paracetamol. Non prescribed products could be used to say that it could be interesting that they have those products as an upsell.

**Is there a lot of unavailable data?**

Data and time is useful for customers. The demography is also useful like the zip codes used.

**When working on forecasting is there a time frame you're interested in?**

Not really. But out of experience if you do things with time frames, do it on the short term.

**Are you interested in the current trends related to covid?**

Yes but it would be difficult to make. There might be a problem with things like the flu, so what we could look at is the impact of flu in the different seasons.

**Are there any external factors that influence the business like weather?**

Yes, they think there are some things that have an influence like climates and seasons. for example heat waves. Demography could also have an influence on the business and health situation of the clients.

**The 11th column of the dataset, is the price of the order also the price of the shipping costs?.**

It's the price of the product only and there is no shipping cost involved. Each row also has one product involved in the dataset.

**What are the major concerns?**

The health costs are rising and the amount of money people want to spend is being minimized. For the pharmacist as a whole it would be nice to see how the product could become cheaper. Others things could be to make atc code used more for prescriptions so that the pharmacist could have more information available. So it could make sense to detect where they lose money while providing more expensive medicine. Another thing to look at is to combine orders that detect patients. So that you could see the time when the medicine has to be given or which medicine could be used together etc to limit the cost.

**What has been done so far for the optimization?**

What they hope to achieve with the data is to bring out stuff driven by data that their customer has overlooked.

**How accurate is the data?**

It is accurate because it's real data from pharmacies.

**Could you elaborate the case? What do you expect from us?**

What they expect from it is that we can show where the money is in the data. Where could the organisations improve, like making agreements and predictions in buying stuff and how to increase the upsell. We could also look at where the pharmacists are giving advice about. Which pharmacists are underachieving compared to others. Another thing could be for example to look at why some pharmacies sell less than other pharmacies. To be aware of while benchmarking, is not to benchmark a pharmacy with one pharmacy but with a group of pharmacies.

**Will they provide other services next to the datasets?**

They will look at our business proposal and give feedback about our business proposal. If there is a need for it we can ask extra questions later in the project. A conclusion could also be that there are no correlations between data. This could also be some kind of insight for the customer.

**How will we be evaluated?**

Think out of the box and don't focus on one thing in the project. Try to focus on multiple things. This way you could find more correlations between things in the data. The quality of the report is also important, so not how big it is but how the report looks. Like spelling errors, page numbers, a background about the people who made the reports.

**What challenge would we face while using the dataset?**

The first one is to understand the data, get a little bit of subject expertise, so we know what the data means. The second one is data quality, so look at what data is wrong and what is missing in the dataset like values or does the missing value mean anything. The most difficult thing to find is if there are missing records. It would also be a challenge to think about the data as if you're the customer. And look at the topics you would like to investigate because there is limited time for the project.

**What methodology do you use?**

Methodologies are difficult to choose, but we use crisp dm(?). Since last year they used the data ops methodology.

**What is the objective of the project?**

Give insight, show how the pharmacy could make profit or lose money for buying and selling their products. Try to find the highest profit margin for the pharmacy customer. The pharmacy wants to make the health situation in Belgium still possible, this can only be done by cutting down the costs in health medication. They also want to keep the health affordable and keep the patients as healthy as possible. So if we could find something that is healthy for the patients it's a good find for the project.

**Can we combine the dataset with other datasets to make a comparison?**

Yes you could combine the data, but the chances of the same datasets are very small. What could help is to combine the data with other datasets that could support the analysis rather than looking for the same kind of data.

**Are there data driven measures applied to analyse the data to get better in their field?**

They normally do it for customers so it is difficult to compare with their own company. What they did learn is that there is not an algorithm that fits every question made by their customers.

**Can we contact Rob when we have a question?**

Yes, you could do this by mail. We can do this directly ourselves. But what rob would like to recieve is one person that is dedicated as contact person per group. The second thing Rob would like is to specify our class and group id so that he knows which group he is talking to.

**Contact Person:**

Email address Rob Prop: r.prop@informa.nl

# 2.Conversation: Danny & Simona - 9/17/2020

## 2.1.Notes

* Base our choices on the data that we received. Make correlations, timelines etc. With the findings in the data we could make a business proposal.
* Do exploratory **data analysis** as a starting point. Look at things that stand out in the analysis. Make this the plan for this week.
* Try to come with a concept business proposal within the coming 2 weeks.
* The proposal should say that we are going to research something that could be beneficial to the company.
* If something in our proposal doesn't work out we could say that it doesn't work for the company at the moment. It doesn't have to be worthless because you could still make a statement about it.
* We don't have to make a final choice yet.

# 3.Conversation: Danny & Simona - 10/01/2020

## 3.1.Notes

* Do the machine learning exercises, project en challenges. These are the priorities.
* For organizational content it's better to do the project and challenges.