alexmaruta [Follow]

doer. thinker. gin drinker. traveller. www.amaruta.tech

Oct 26, 2017 · 3 min read

Data Analysts, Scientists and Engineers

I often find myself explaining what is that I do. Over time, telling this story in cabs, elevators, or in the train (yes, I travel a lot for work) I perfected my approach and now I think I am in a place where most people "get it". Compared to a couple of years back where the most I got from them was a ... "sounds cool" reaction now people start to have a more sensible understanding. Anyway, I thought I should put this in writing just to keep as a reference for the future and maybe to help others choose a career.

The way I always start this conversation is by explaining that i) I work with data and ii) within the " data space" there are mainly three different roles:

1. Data Engineers

2. Data Analysts

3. Data Scientists

Then I quickly tell people a bit about each of the three roles:

- **Data engineers** are those who manage the data infrastructure layer (in simple words, the databases), or are busy deploying models or automating data processes (ETL). These guys do the heavy querying, creating data models, architecting databases, etc etc. Or if you prefer an analogy, think about building the house, these guys would design it and build it.

- **Data analysts** are those who develop insights by analyzing the data. They do exploratory data analysis, data quality checks, data cleansing and ultimately define business questions and try to answer these by telling a data "story". The story is visualising the data that is why this "story telling" capability is essential for the analysts. Based on the story, insights are created that support

business decisions. To refer back to the house analogy, these guys would analyze the house and come up with insights like "house should be rented, or sold, or put on airbnb, or used as a commercial space" etc etc, anything that would help drive better *business decisions*.

Data scientists are the champions. These guys build statistical/predictive models that once deployed in production generate new insights on the fly. In terms of the house analogy, these guys would predict macroeconomic factors and their impact on the house price.To make it even more simple, the scientists build models (forecasting, machine learning, etc) and the analysts consume existing data.

Now to go back to the question "what is it that I do", I always say that I am right at the crossing between these roles:

I do loads of data preparation which means querying the databases, extracting data, cleansing it, preparing it for analysis (step1).

Then I come up with hypotheses/business questions. To do this I use visualization tools to analyze data and get more understanding of trends, patterns, correlations, etc which I share with the business (step2).

The final step is to make some statistical inference, maybe even build some models (linear regressions or decision trees mostly) to predict future behavior (step3) and push these into production however what happens in practice is that by the time step1 and step2 are completed there is little budget left :) (joke alert!)

Or most ot the times I just say "I'm in IT." That pretty much simplifies the conversation a lot…