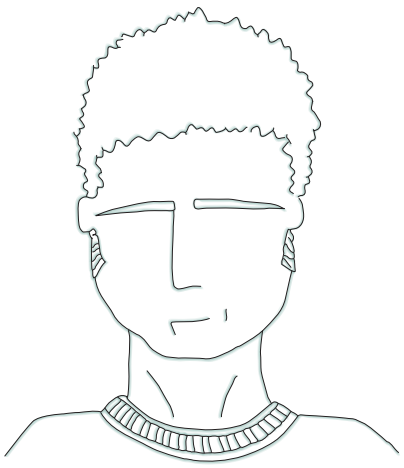
RESEARCHER AT THE UNIVERSITY

Marco Lancia, Eng, PhD

'I spend 30 minutes early in the morning each day to provide answers and support in online community forum. It is like a competition'



Age: 34

Occupation: Post-Doc Researcher

Organization: Leiden University Medical Center - NL

Marco **appoints himself as novice Data Scientist since 6 months**. He got a Post-Doc position at Leiden University Medical Center and works in Survival Analysis in Oncology.
Marco gained a Master degree in Engineering and a PhD in Mathematics and didn't own any specific knowledge in the oncology domain, being his previous experience in a research project in the transport sector.

When Marco decided to apply for this "informal" Data Scientist job position he only knew that **he wanted to improve people life quality with his expertise**. It was a real chance to change his research work paradigm: **from a purely theroretical and modeling work to application to data analysis and looking for evidences in data.**

Marco **has now a pivotal role in the research group**: he does data science as well as model implementation and data interpretation to share information with medical doctors. His data analysis is useful to understand specific phenomena in oncology.

Thanks to these experiences he understood that a **formal preparation and recognition as Data Scientist could be interesting** to easily join future projects in different domains.

Key Characteristics

- He believes that the DS provides information used to take decisions
- He studied the literature as self-learning and didn't receive any formal and structured training on the domain knowledge
- He refers every day to online communities to look for updates/ resources.

Questions

- Is there an entity that release official DS certification?
- How can I overcome my gaps in specifics competences (theroretical and practical) to become an expert DS?
- Where can I share my work results resources and discussion for other people interested in data?

Goals

- Gaining a DS certification
- Being recognised as an expert from the scientific community
- Mastering DS skills to apply in several domains (not only in the medical field)

Needs

- Getting support from the community members with contributions, experiences, references and discussions
- Getting in touch with data experts and looking for expert collaborators (also through direct meeting)
- Get positive evaluation from community members
- Practicing with real data sets

Influencers

- Giving a real support to several field of research
- Personal management of activities
- Being a part of an active community of DS experts
- Gaining money from a recognised title as DS
- Operating in several domains with other new collaborators

Frustractions & Pain Points

- He doesn't feel a real DS: "I feel I am now more a statistician"
- He needs increase his education in DS domain
- He mind the risks to get his data / dataset stolen: very often data exchange expose researchers to steal of data.

'This is not the Big Data era, rather the Bad Data era. I spent 90% of my work time in cleaning the data and preparing the data for the analysis'

Age: 42

Position: Data Scientist

Organization: WeLikeCRM - UK - IT

Luca is a passionate statistician and IT specialist in a data analysis company. He works as Chief Data Scientist and **develops solutions that transform data, via statistical modeling and machine learning procedures, into actionable insight to enable business decisions and optimize operational processes**.

The **most important requirement to become a Data Scientist for him is to own a "data culture"**, meaning knowing data, regarding data and treating data according to a scientific approach.

Luca is **pretty familiar with online courses and online DS community of big data practitioners**: he is a member of Microsoft Azure community and usually receives solicitations to provide the community with questions and input to online discussion.

Luca's approach is focused on undestanding how the scientific method is applied in real situations to look for meanful insights. He would like to have formal and official resources, gain the DS certificate and be regularly updated.

Key Characteristics

- **He is currently working as an indipendent practitioner in: descriptive analysis; data statistic exploration; data interpretation for data-driven decision**
- He is building a "data culture" in her company
- Community, platform, newsletter and blog joined: S2DS, Microsoft Azure, Stack Over Flow, Github, R-ante, Insula-R, DS LinkedIn groups, Data Science Weekly Newsletter
- He provides support to novice users

Questions

- How do other DS practitioners formalize their competences?
- How can I increase my personal website in order to be recognized as expert?
- **Is there a multi-level qualified certification in DS?**
- Where can I meet the most influential experts in DS to work and practise with?

Goals

- Gaining a DS certification
- Gaining higher professional development levels
- **Develop my career on regularly basis**

Needs

- **Practicing with real datasets in DS team work**
- Attending events, meeting and conferences to get in touch with DS stakeholders and find new partners
- Providing the novices with support for practical exercise/ problem solving or developing tutorial and educational materials

Influencers

- Qualified DS certification that provides higher educational and professional development
- **Sentiment and opinion of DS experts and scientist**
- Identifying other fields of application for neurostatistics

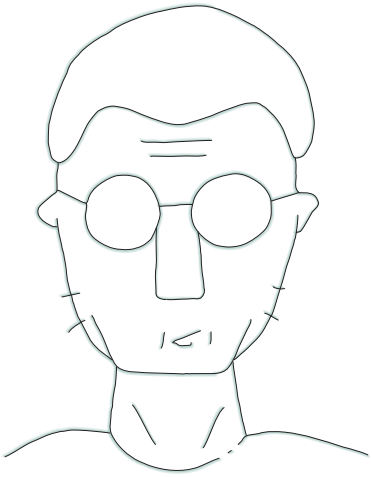
Frustractions & Pain Points

- **No much time to regular active contribution in online communities**
- Finding expert partners to collaborate with
- Finding highly specialized training to reach higher professional development levels and advancements in his career.

CEO AT CHORALLY

Renato Castrovecchio

'I wanted to invest in one of my Data Scientist by making him enroll to the So Big Data Master. I would have paid the fee and allowed him to attend the courses.'



AGE: 43

POSITION: CEO

ORGANIZATION: CHORALLY - UK - IT

Renato is the CEO of a specialized social CRM company, dealing with listening and data analysis, customer engagement and advocacy. Renato needs to enroll **Data Scientists with a very complex set of competences ranging from statistics and modeling to data interpretation, until skills in IT infrastrucures and software development.**

Renato looks for candidates **by targeting conferences talk, blog posts, papers and networks in order to better understand profiles and attitudes.** It's not important be a domain expert since there are not specific tools or best practices that the candidate need to know.

His Data Scientists work autonomously for 70% or 80% of their time and for the rest they **need to collaborate to ther other teams, such as marketing, IT and design**. For me is essential that all potential Data Scientists are able to manage data and **gain valuable information from that.**

He knows that in online communities and platforms there are **people interested in problem solving and in actively joining other projects**: these virtual places could be the right start point to access and meet future expert Data Scientists.

Key Characteristics

- He prefers referring to online places and focusing on members contributions
- **He looks for people able to make data interpretation and which are interested in problem solving**
- Needing to enroll DSs owning soft skills such as communication, negotiation and accounting

Questions

- Where can i find the DS profiles that I'm looking for?
- **Is there a list of DS user ranking?**
- Can I choose DS candidates from a list of users competences and experiences?
- Are there instituions or universities that could qualify young candidates able to join my company?

Goals

- Building a team of DSs
- **Finding candidates with specific skills and competences: IT and statistical knowledges; degree and background; previous experience in DS project**
- Need to own the ability to explain, discuss and disseminate contents, methods and results of their work

Needs

- Finding a source of DS practitioners and experts
- Looking for interesting users online contributions
- Getting in touch with possible candidates (e.g. providing a competition) and setting up real meetings
- **Evaluating the members results and recognising them as DS**

Influencers

- **Participating in competitions and users ranking**
- Contributions in term of resources, materials and references in online forum, blog, ect.
- Online communities (Linkedin Groups on DS, Facebook Groups, Google+ Groups, etc)

Frustractions & Pain Points

- **Finding mature DS with a very complex set of competences related to data management, analysis and interpretation**
- Too much time in searching contributions and material in several communities or forums