

DATA SCIENCE

TRAINING PROGRAM

Lab 1: Data Science Discussion

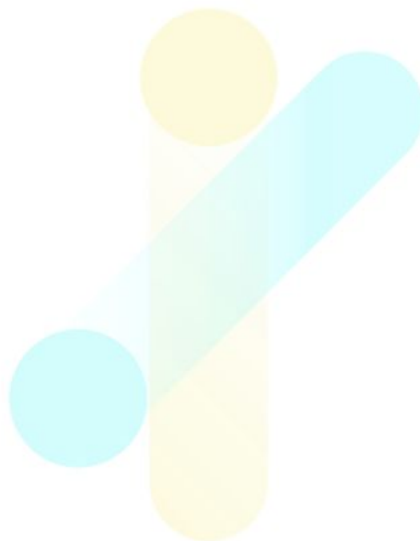
Nico Rotstein, Kodit.io
2019-10-09

Labs



Practicalities

- Accounts needed:
 - Google Colab for notebooks: <https://colab.research.google.com>
 - Kaggle for datasets: <https://www.kaggle.com>
- How to lab:
 - Lab projects: dataset + challenge (no tests!)
 - Groups of 1-3 people
 - Get staff help on Mondays after lecture, and Wednesdays
 - Questions to peers on the Community Platform [www.theshortcutcommunity.org]



Certification Project

- Choose a dataset
- Apply all concepts learnt in the course
- Groups of 1-3
- Presentation on the last week



Data Science Fire Starter

- **Startup:** automated complaint resolution
- **Pilot:** internet service provider
- **Dataset:** tens of thousands of complaints
- **Features:**
 - complaint as free text
 - complaint category
 - amount requested
 - amount refunded

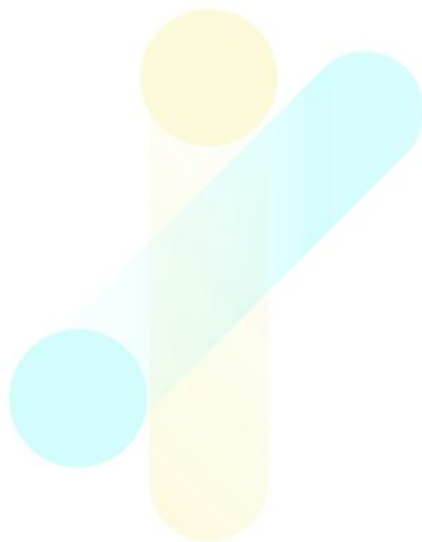


Dataset

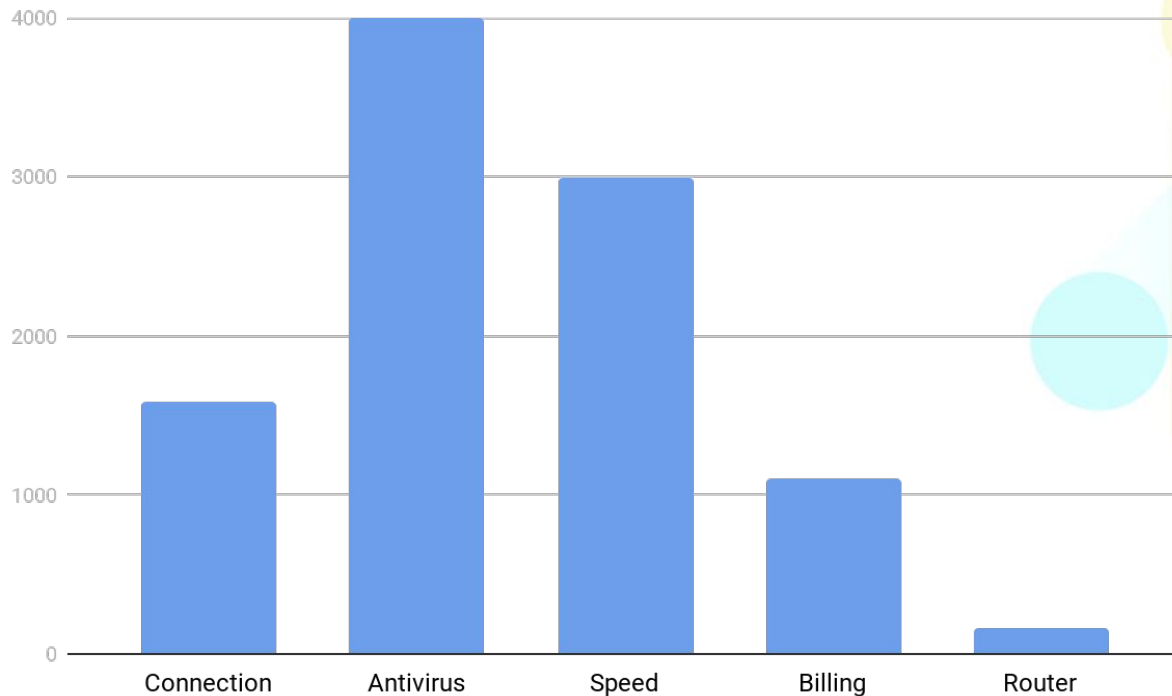
Complaint	Category	Requested	Refunded
<i>I never requested the antivirus package</i>	Antivirus	100	50
<i>Please refund the antivirus charges</i>	Antivirus	0	25
<i>My connection is too slow</i>	Speed	30	60
<i>I ordered 100 mbps, not 100 kbps</i>	Speed	30	30
<i>Doesn't work; want my money back</i>	Connection	75	0
...

The Questions

- What's the problem we are trying to solve?
- What can we squeeze out of free text?
- Compromises, anyone?
- What do we want to shipping first?



Complaint Categories



Your Articles

