Data Related Roles in todays Businesses

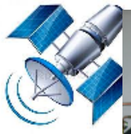
What possibilities are there after your PhD!

@Piotr Laczkowski

piotr.laczkowski@gmail.com

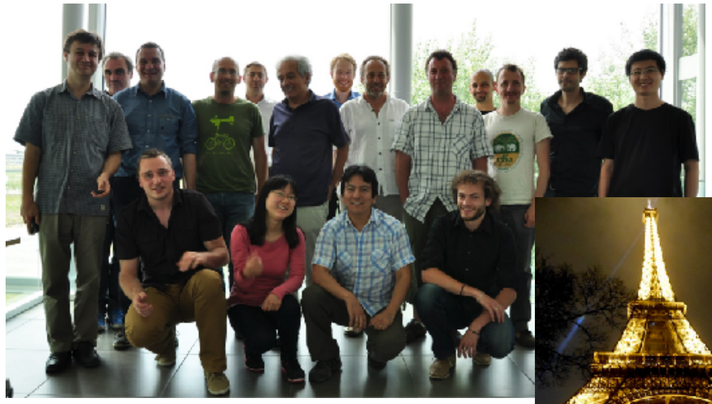
Polska
(Poznan)

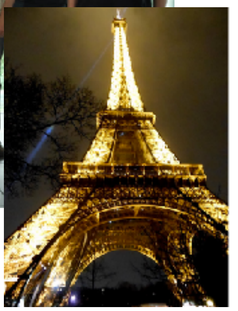
Grenoble
2005



PhD (CEA)
2009-2012

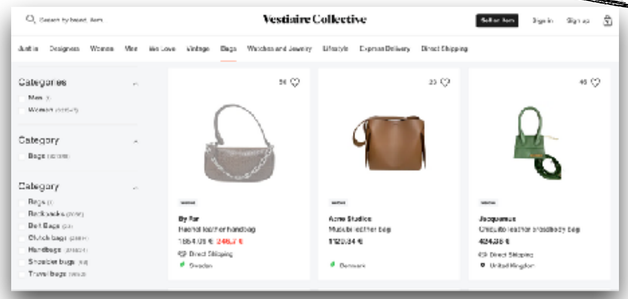
PAris
2012





back market
2016

Vestiaire Collective
2019



LV +JPA :)

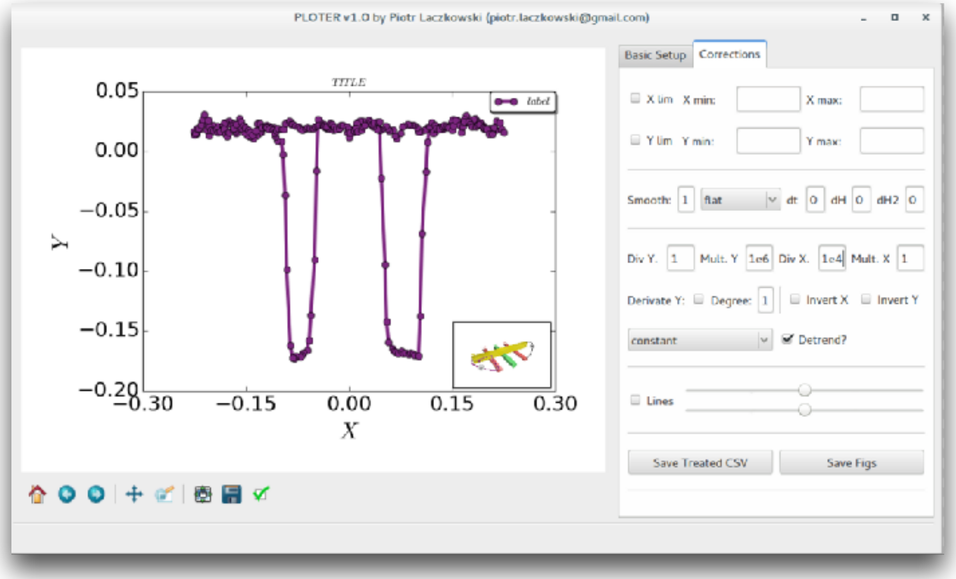
(1) Your are THE ONLY ONE responsible for your path !!!
-> think about it already (plan it)

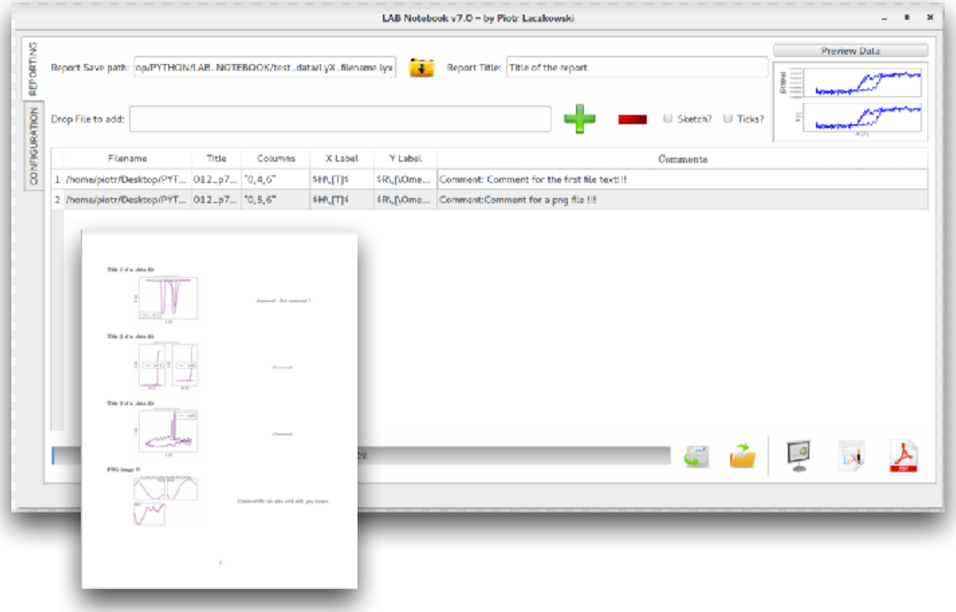
(2) use this time (PhD, ...) to discover what do you like to do
-> HAVE FUN doing it and gain skills !

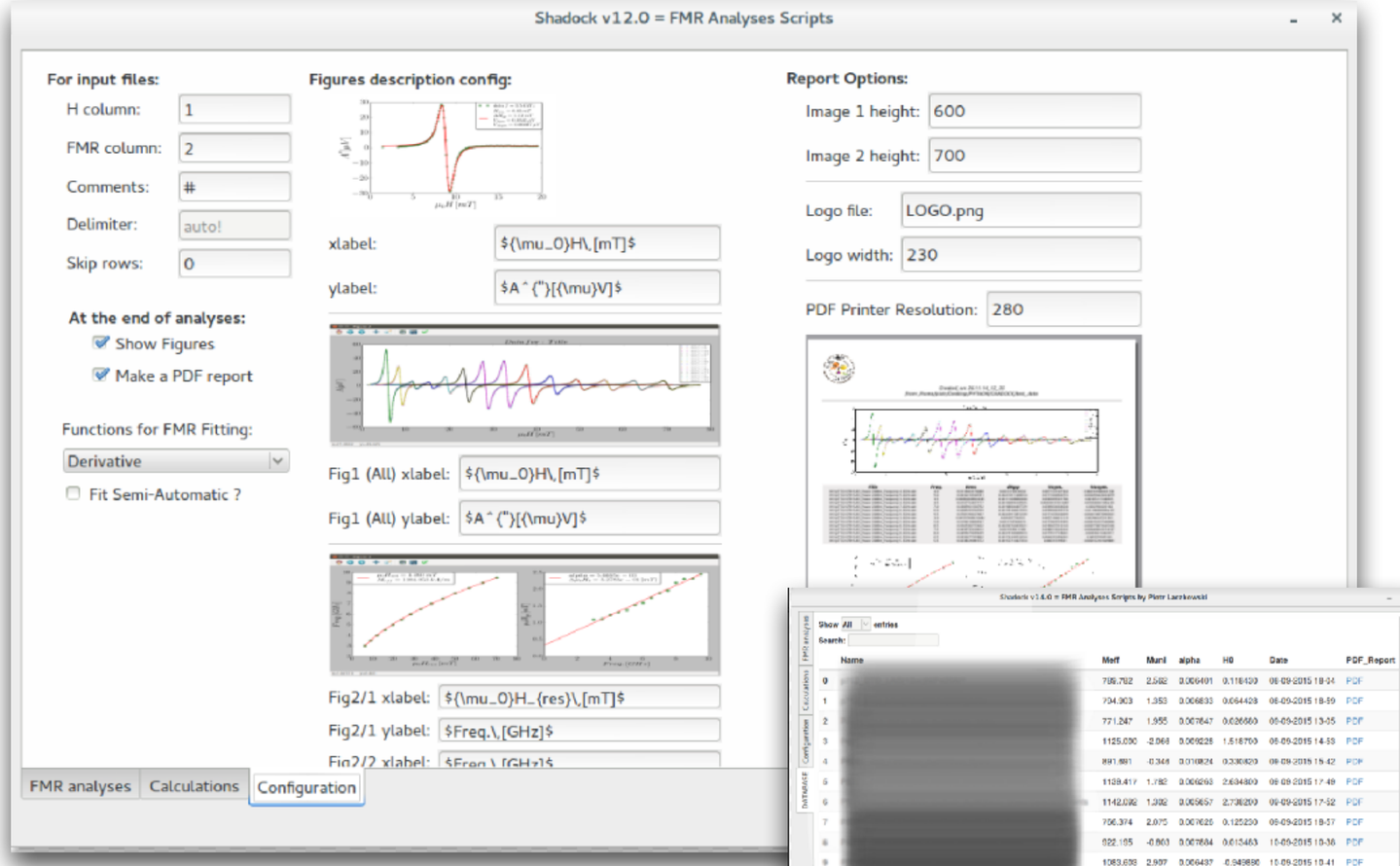
(3) If you have skills -> you DO NOT NEED any network not luck to find the job, it can find you instead!

PhD + Post-Doc
Side Projects

Examples of Python side projects

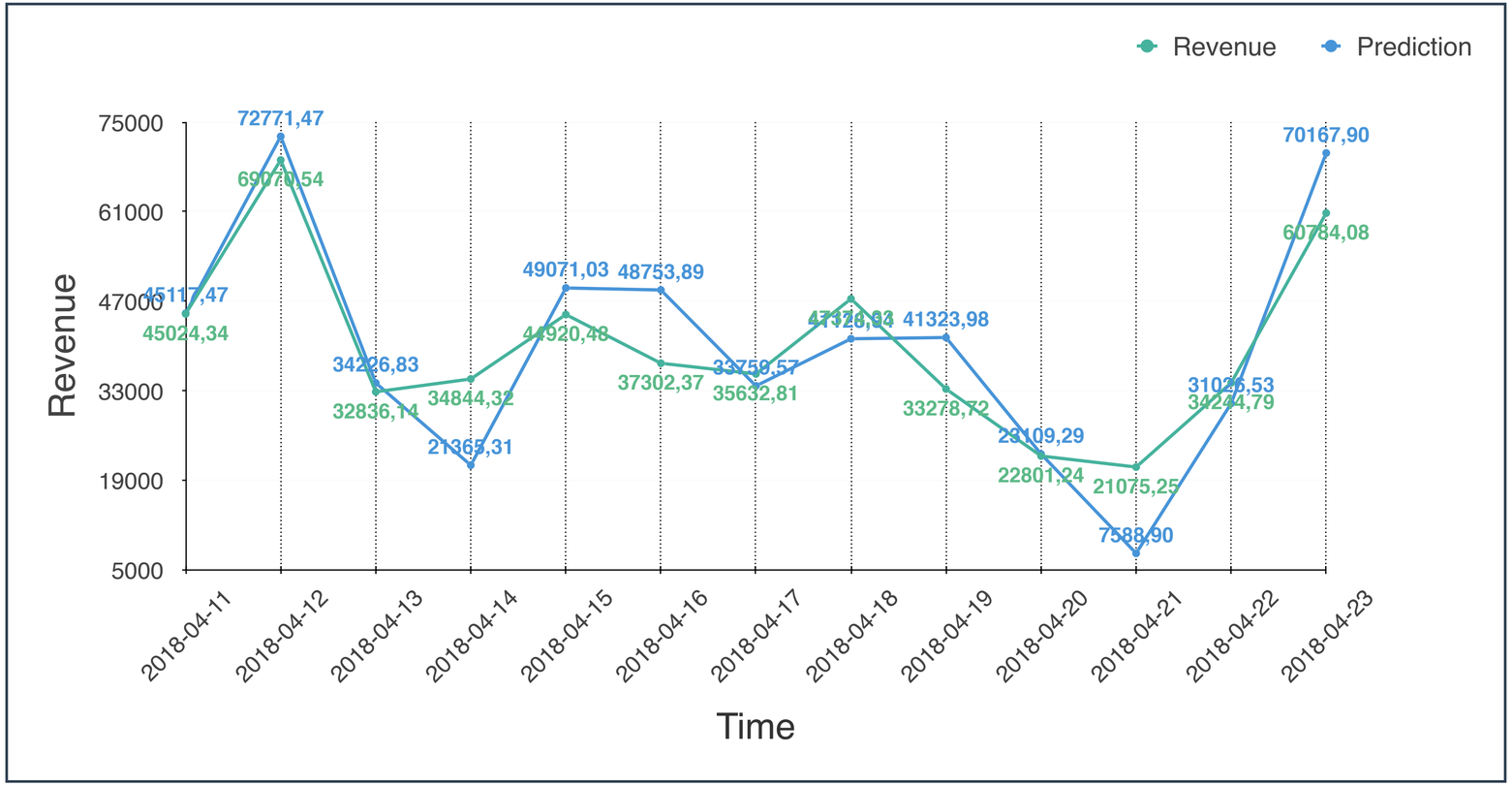






Back Market Projects

Predicting revenues (GMV) based on weather



Predicting GMV while looking if its rains or shines with accuracy > 80%

Clients Alerts Matching system

Iphone 6s 34 go or rose

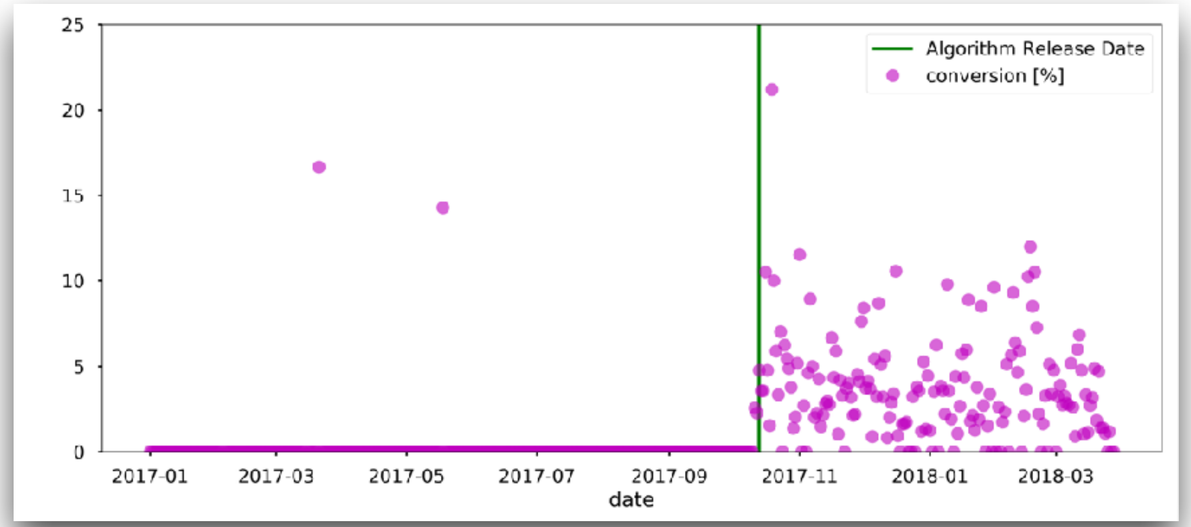
Je veux ce telefon vers 100€ qui soit tout de dant et qui soit or rose 34 go

Huawei p8 max

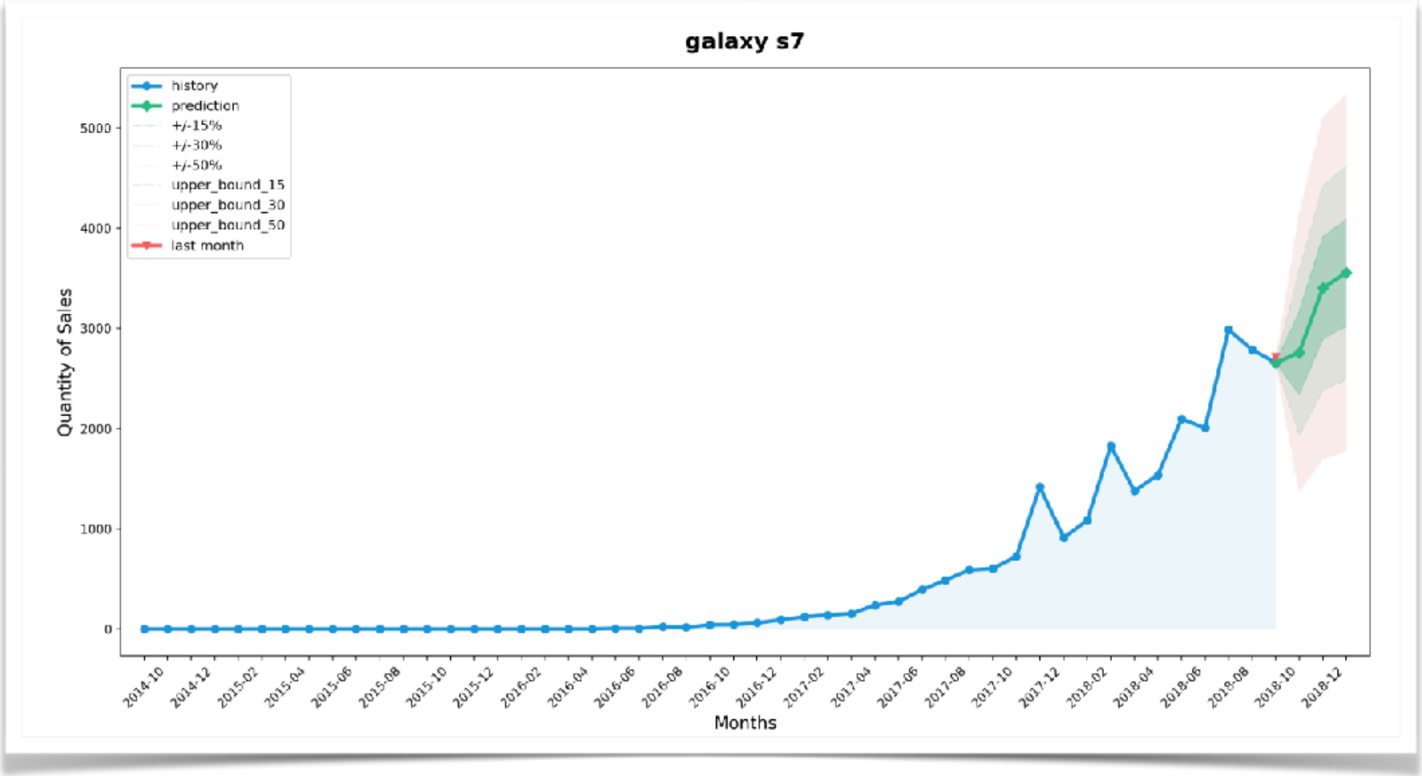
Version Gold 64go avec son écran de 6.8 c'est idéal pour tous les medias et même le gps se téléphone et au top dans tous les domaines je regrette de m'en être séparé (débloqué tout opérateur) les derniers temps on le trouver à 229€ neuf donc grand maximum 300€ (il étais vendu avec so protection cover dans sa boîte)

Machine Learning

Increasing Conversion rate !



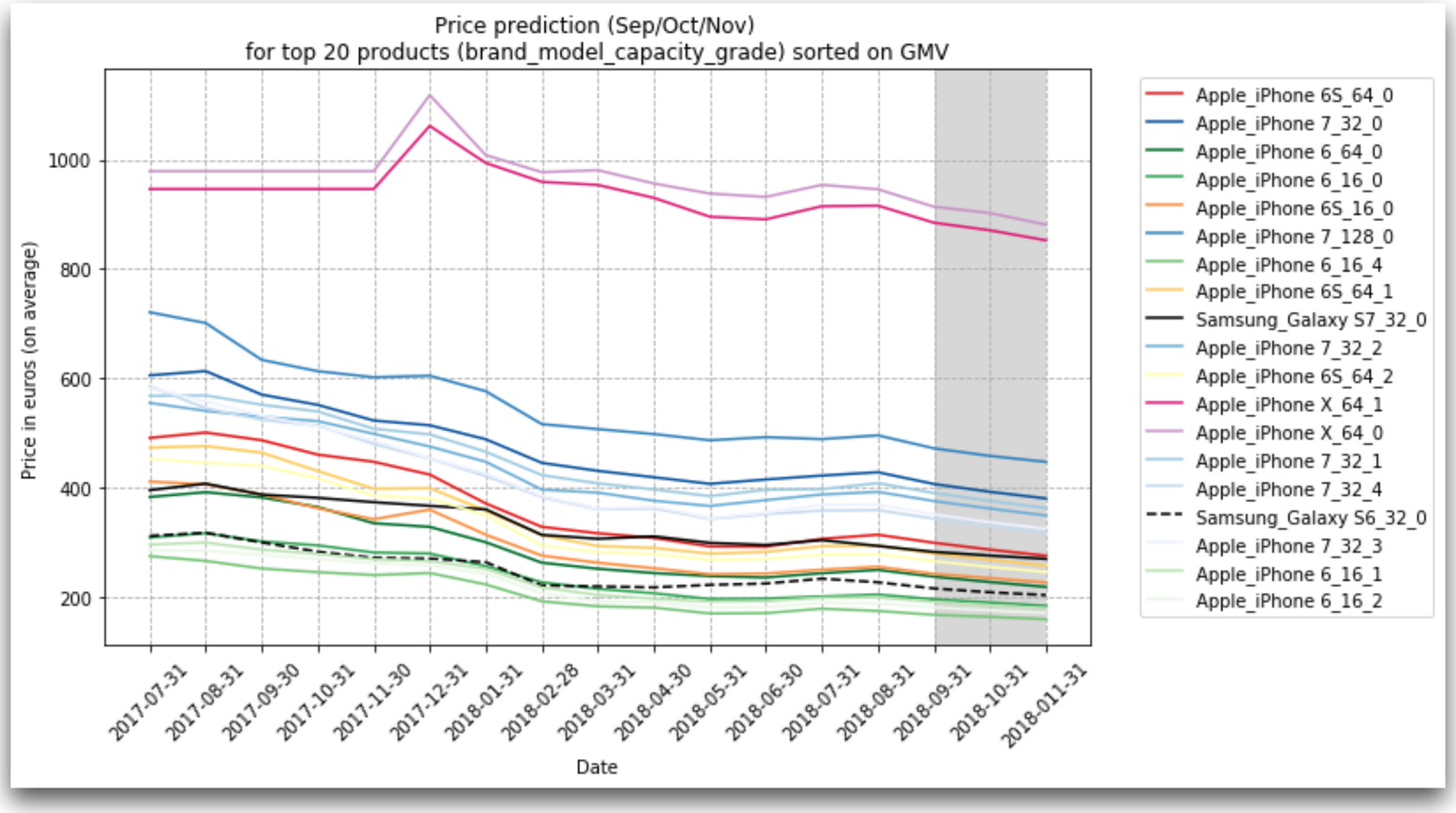
Products sales perditions - supply strategy planning!



85% of the GMV covered with error < 5%

95% trends perfectly predicted + **85% GMV Covered**

Prices Predictions

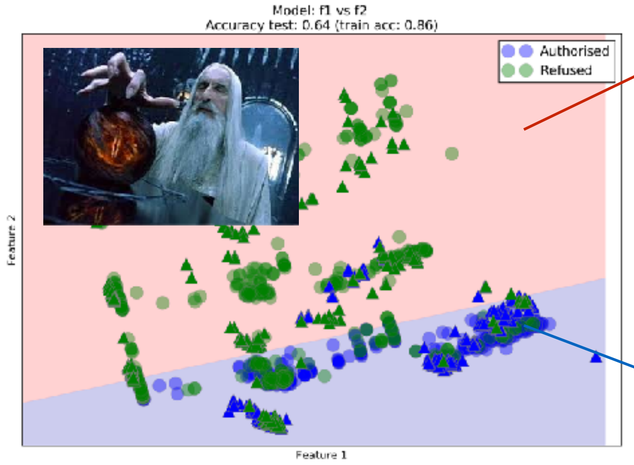


85% of the GMV covered with error < 5%

95.58% GMV trends correctly predicted + **80%** GMV covered

Fraud detection

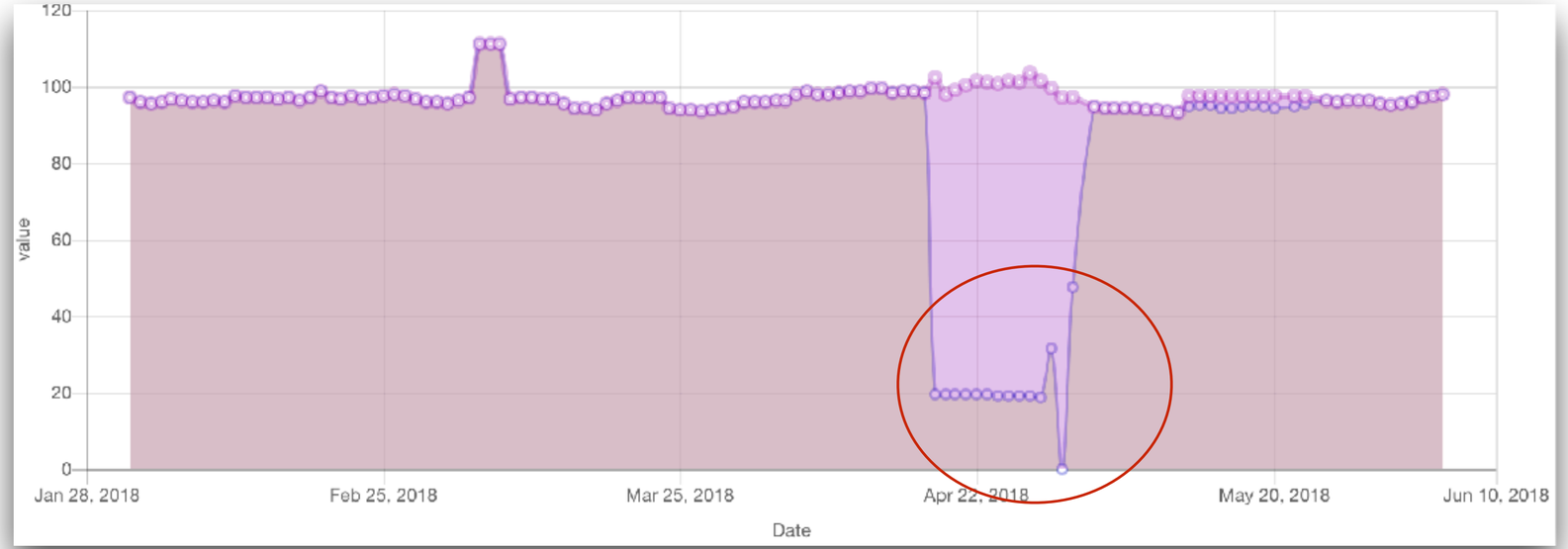
Machine
Learning



ACCURACY: 97%

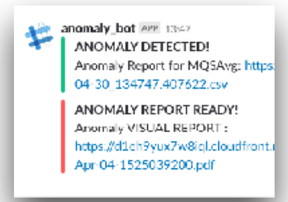
Preventing Fraud !

Anomalies Detection System

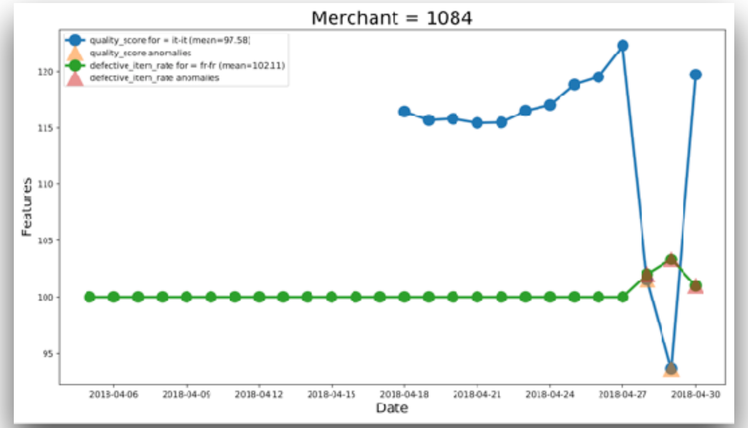


Machine Learning

Slack

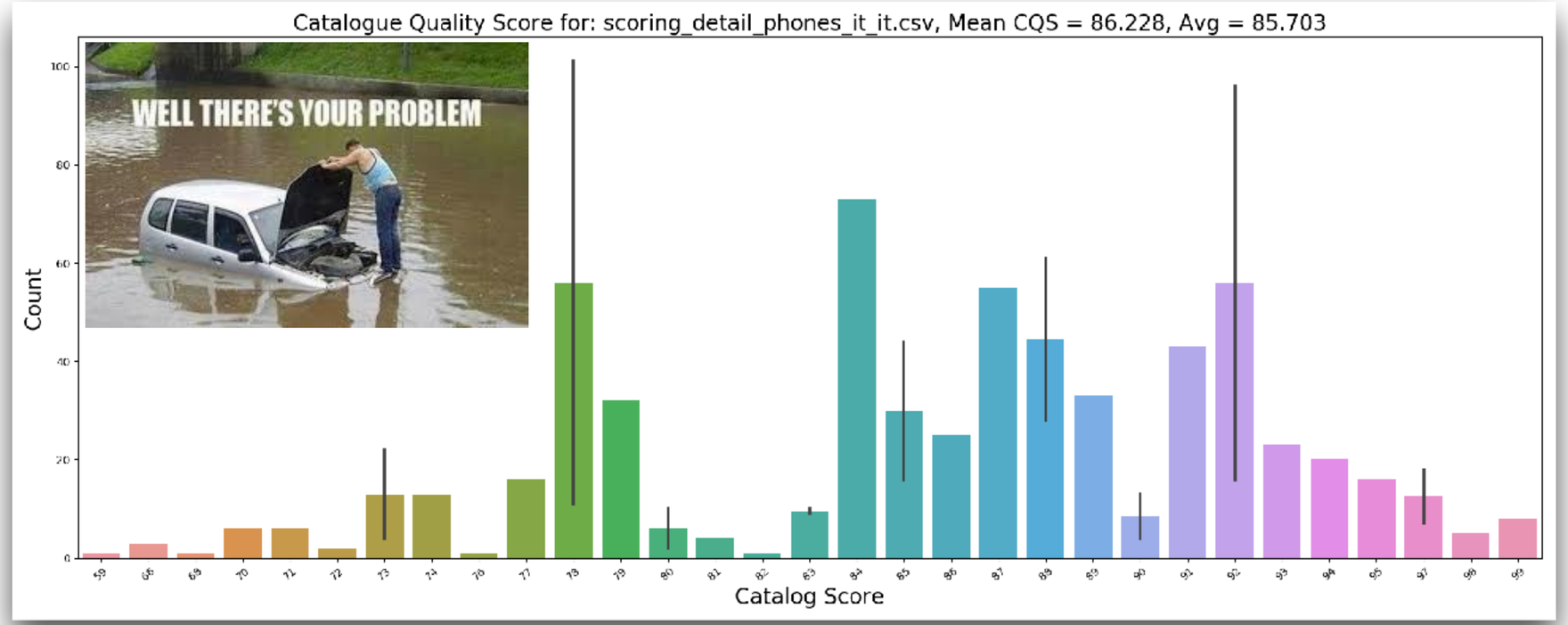


Visual Reporting



Detecting Problems Automatically !

Catalogue Quality Score

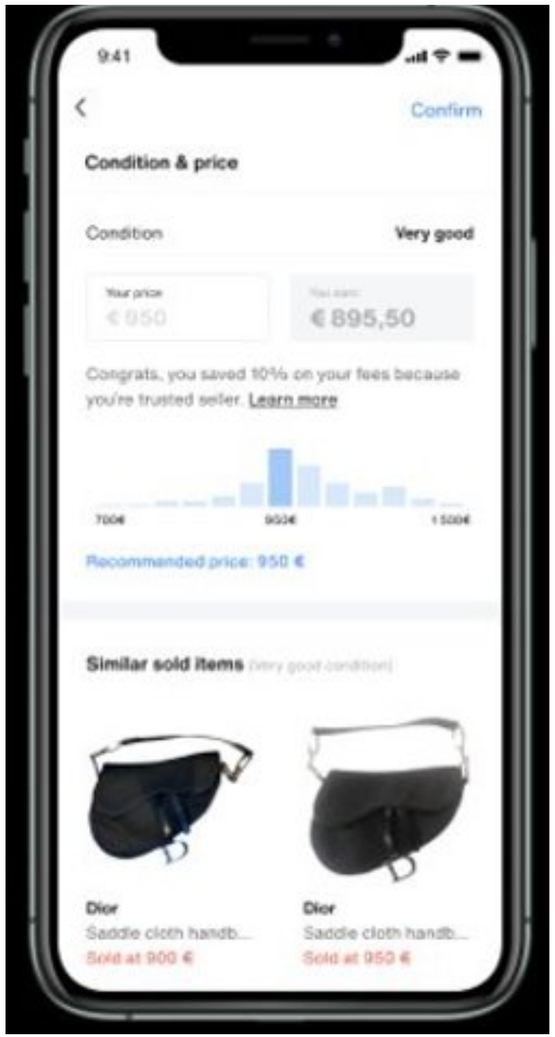


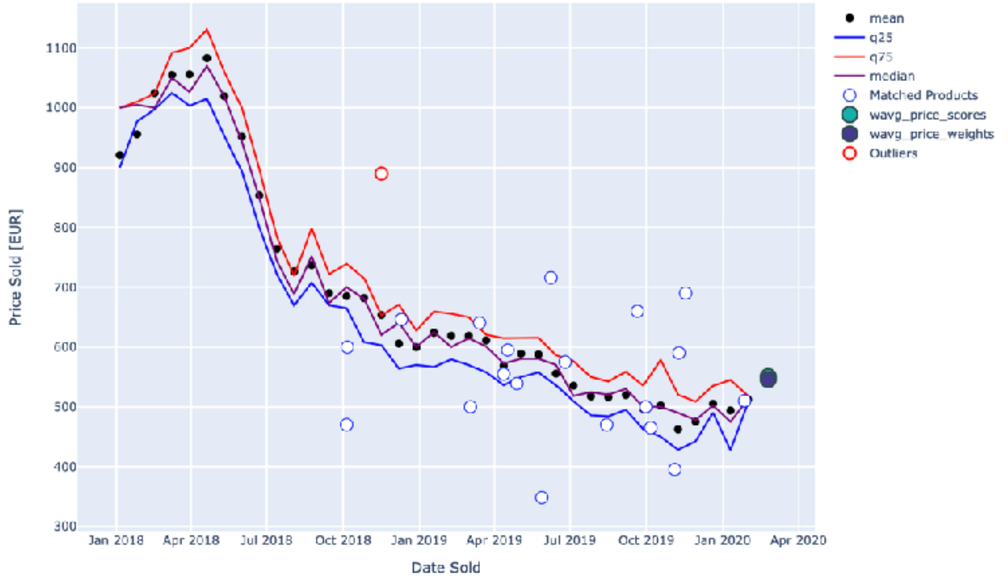
product_id	priority	score	title	color	connector	contrast	date_release	double_sim	image1_url	image2_url	image3_url	ima
3207	6.999140589454527	92.25352112676056	iPhone 6S 64 GB - Oro Rosa - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	8.0	6.0	4.0
3206	6.657680361015061	79.5774647887324	iPhone 6S 16 GB - Oro Rosa - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	0.0	0.0	0.0
1859	5.025024012941712	92.25352112676056	iPhone 6 64 GB - Grigio Siderale - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	8.0	6.0	4.0
15023	4.13774834437086	88.02816901408451	iPhone 7 128 GB - Nero - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	8.0	0.0	4.0
2225	2.8071562716220226	94.36619718309859	Samsung Galaxy Note 4 32 GB - nero - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	8.0	6.0	4.0
3215	2.512512006470856	92.25352112676056	iPhone 6S 64 GB - Grigio Siderale - sbloccato da tutti gli operatori	9.0	1.0	1.0	8.0	0.0	10.0	8.0	6.0	4.0

Controlling/Increasing Data Quality !

Vestiaire Collective Projects

Pricing “unique” items

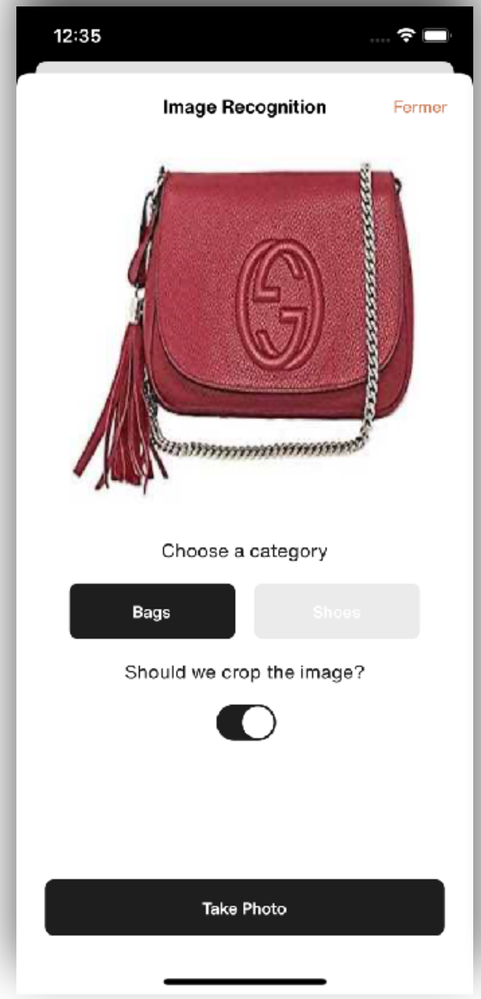


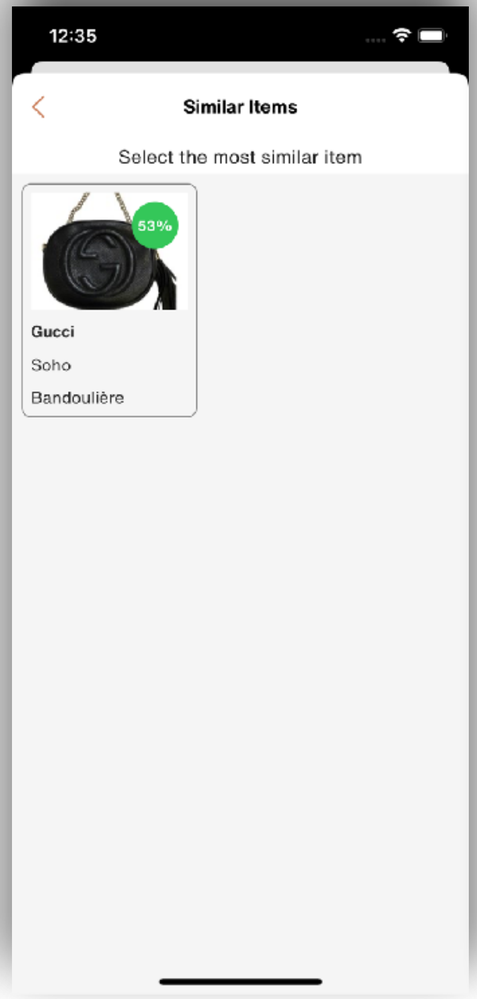


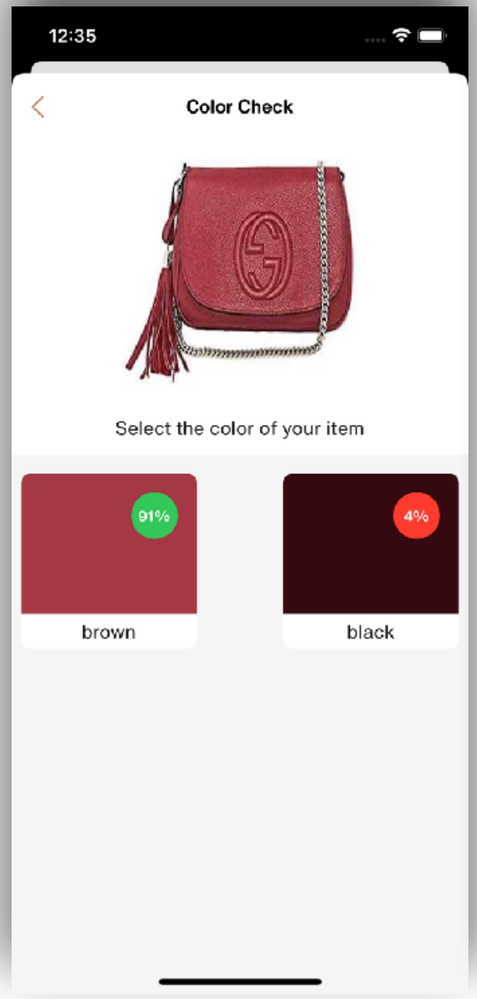
Semantic Image segmentation -> extracting objects!

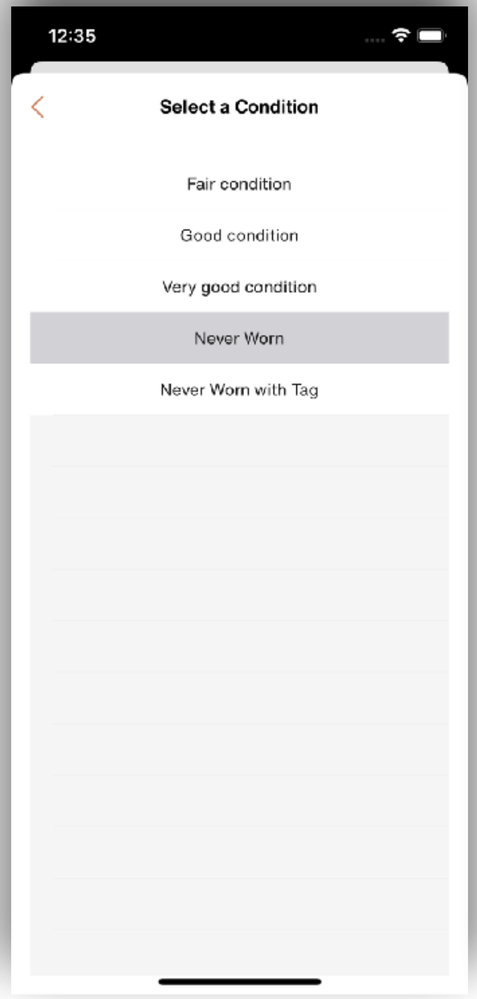
Extracting Colours from the image

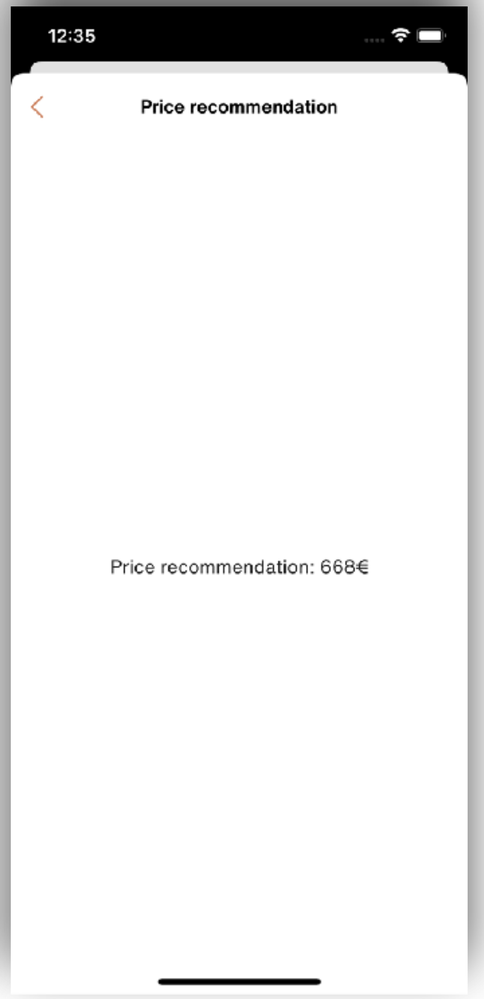
Reverse image search engine





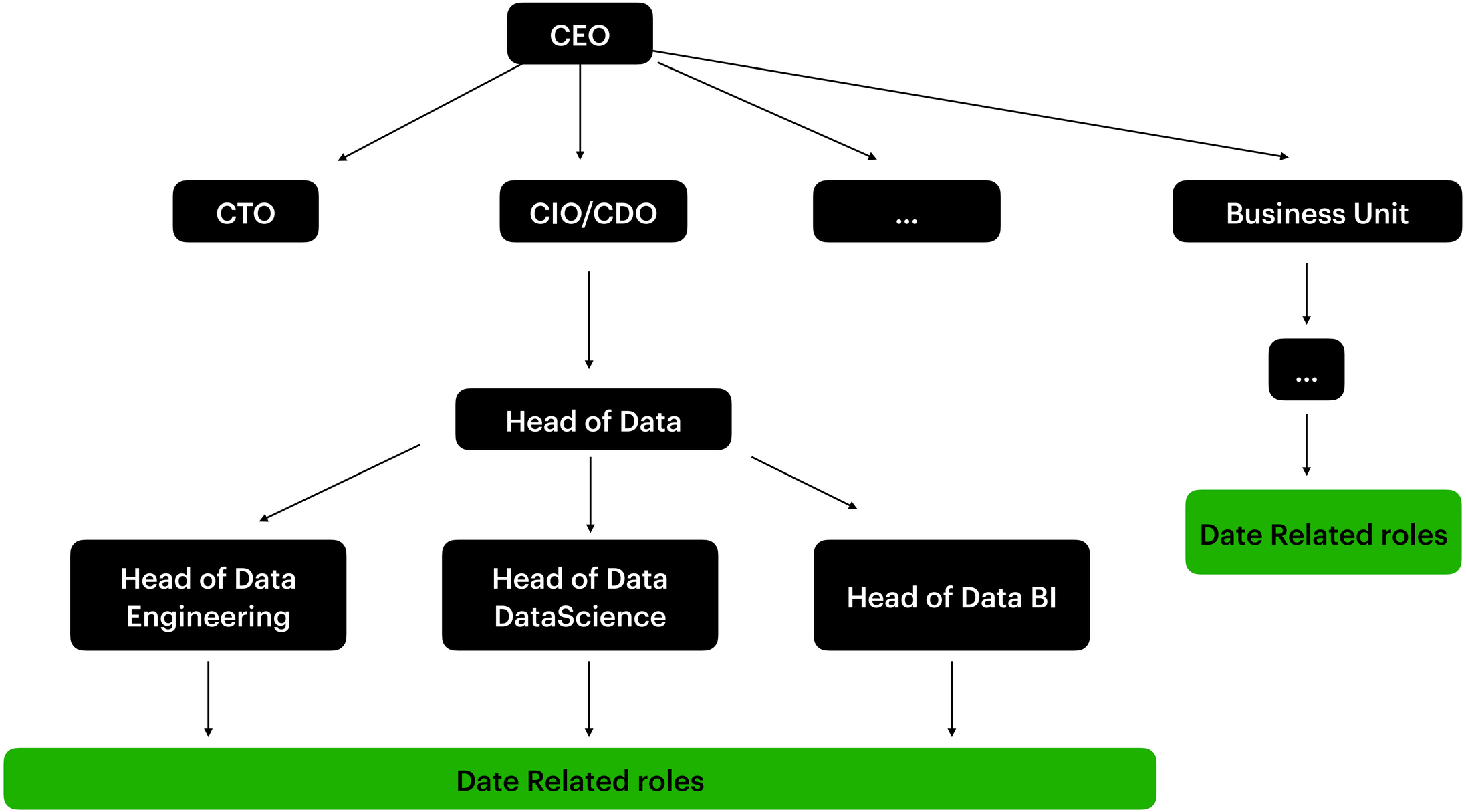






Data Related Roles

Data Oriented Roles in Business - typical tech organisation



Data Oriented Roles in Business

Data Analyst

Business Intelligence (BI) Engineer

Data Engineer

Data Scientist

Machine Learning (ML) Engineer

Typical “Data” career progression

Data Analyst

Role:

Analyze the data to respond to business related questions

Preparing reports and visualise data (storytelling)

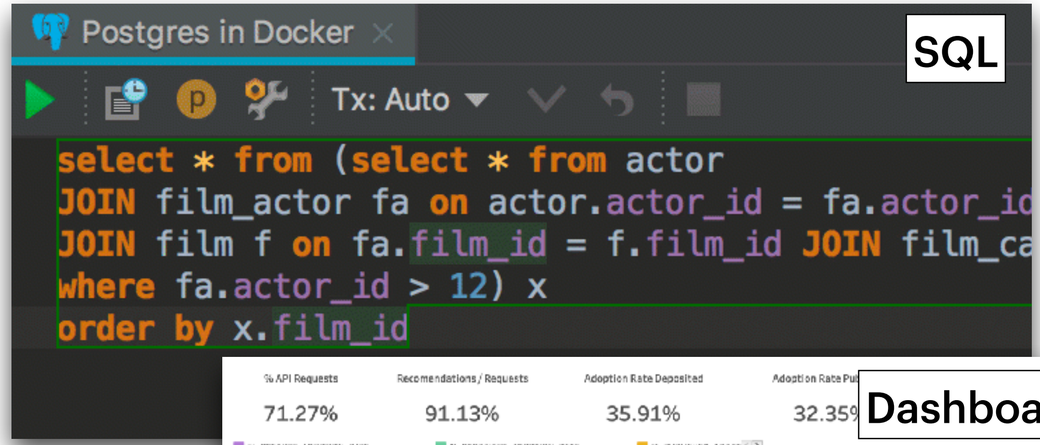
Questions:

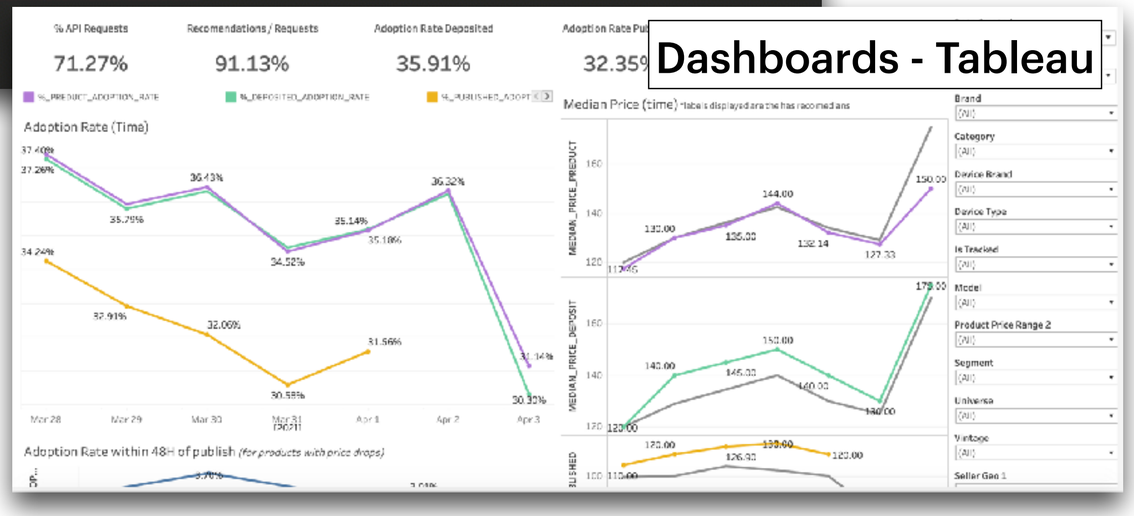
What is the most common client path on the website?

What was the impact of last promotion on sales?

Do clients adopt our price recommendations ?

What do they use (day-to-day):





Tools:

python

tableau

R Programming

SQL

X

Google Sheets

BI Engineer

Role:

Combine various sources of data into aggregated information

Create and provide new data based on existing data sources

Preparing scalable dashboards for business

Questions:

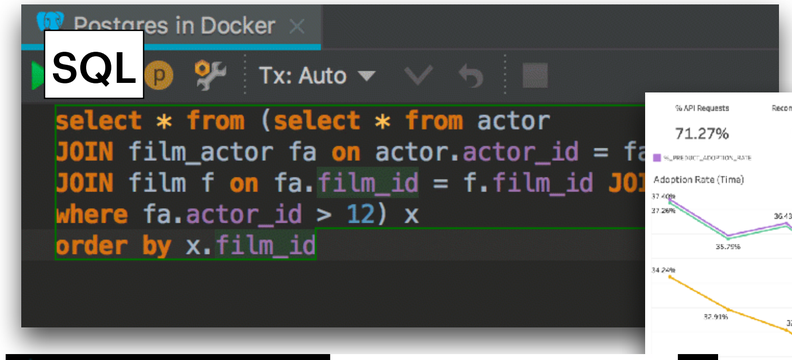
Can you add this information to the product table?

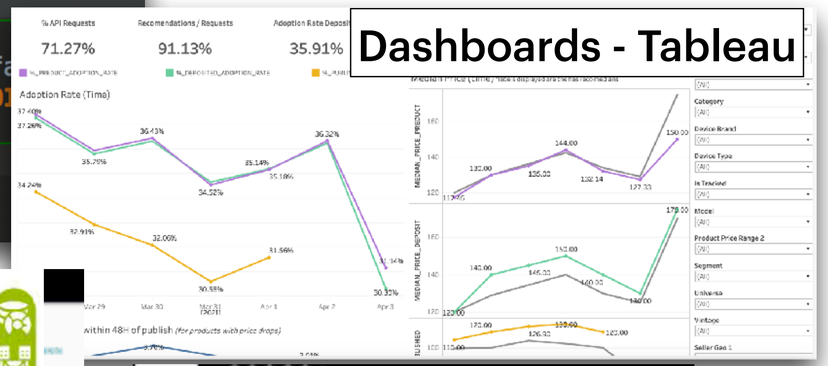
what does this filed in this table means?

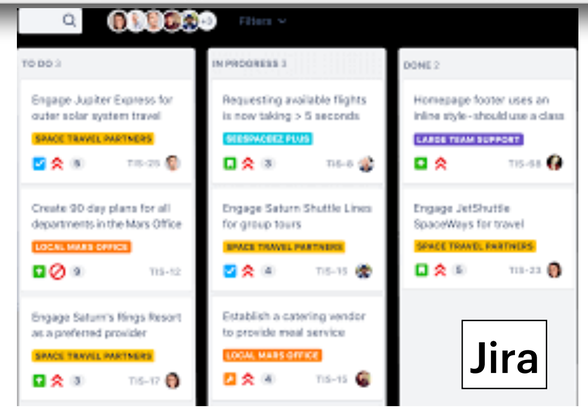
How do we calculate the output price in this table ?

why does my dashboard has old data ?

What do they use (day-to-day):







Tools:

python

SQL

Apache Airflow

MATILLION

R Programming

JIRA

git

dbt_

Data Engineer

Role:

Connect new sourced of data

Manage DataLake (databases, accesses, roles, warehouses)

Prepare data sources interfaces: APIs

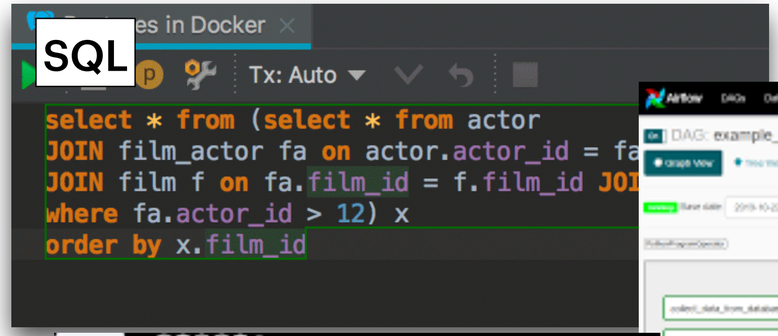
Questions:

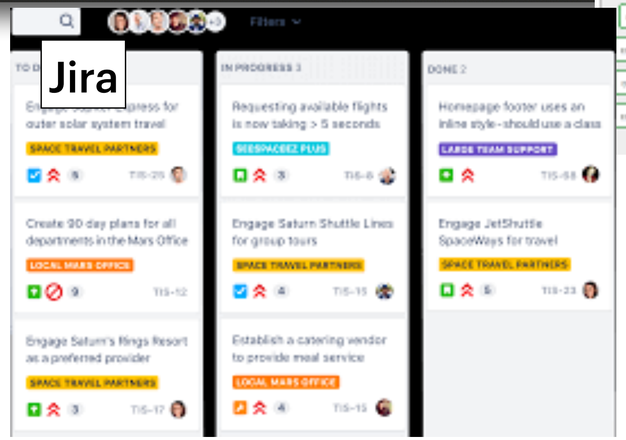
Can you scrape our competitors data?

we need this data to be available for mobile on the API

we need access to this data table

What do they use (day-to-day):

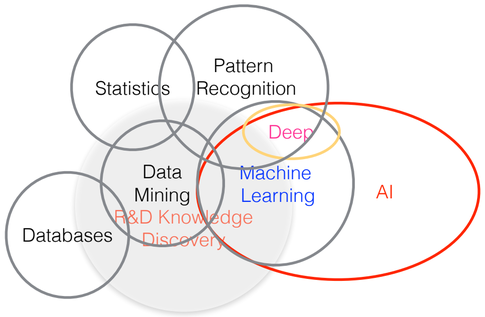




Tools:

python · SQL · Apache Airflow · docker · amazon webservices · Google Cloud Platform · APACHE Spark · R Programming · JIRA · git · FastAPI · elasticsearch · pandas · Jenkins

Data Scientist



Role:

Build statistical models to solve a business problem

analyse and clean the data to select right approach for a problem

manage a dedicated data projects: end-to-end

Questions:

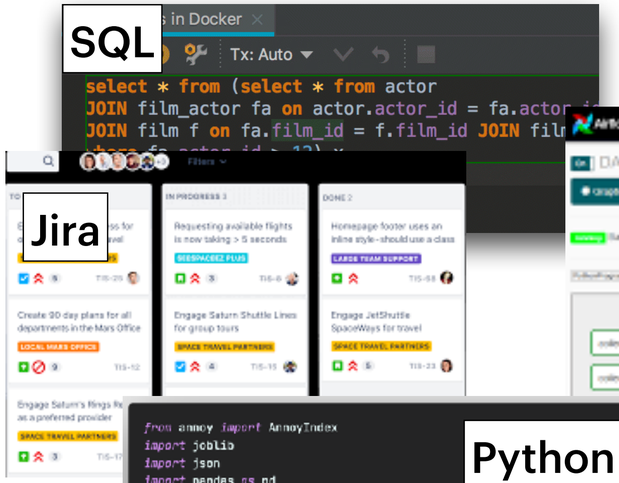
can we predict orders/prices of all products per country?

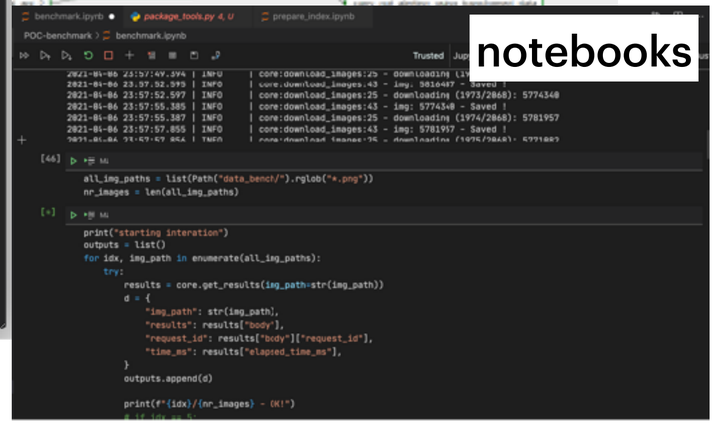
can we fill-in missing values in our attributes?

can we test which feature is better for the user experience statistically?

which products should we show to specific users?

What do they use (day-to-day):





Tools:

python · SQL · Apache Airflow · Apache Spark · scikit learn · docker · amazon web services · Google Cloud Platform · CI CD

R Programming · JIRA · git · pandas · TensorFlow · FastAPI · elasticsearch

ML Engineer

Role:

deploy/scale all data science models technical architecture

provide technical tools for scalability

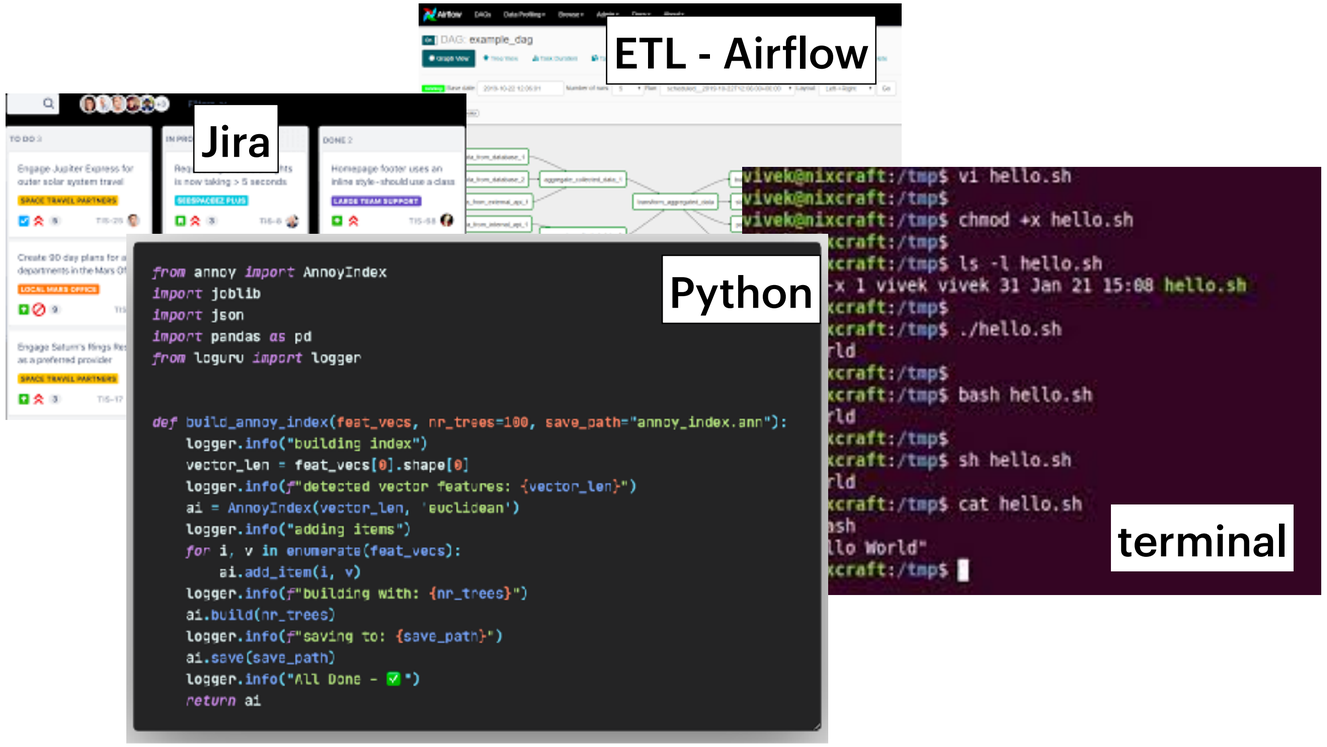
assure best practices in software development

Questions:

how can I deploy this API on Kubernetes?

can you deploy this model to production?

What do they use (day-to-day):



Tools:

python · Apache Airflow · docker · amazon web services · FastAPI · TensorFlow · kubernetes · Jenkins

GO · git · JIRA · Google Cloud Platform · elasticsearch · scikit learn · CI CD · BASH

Personal Recommendations

Recommendations:

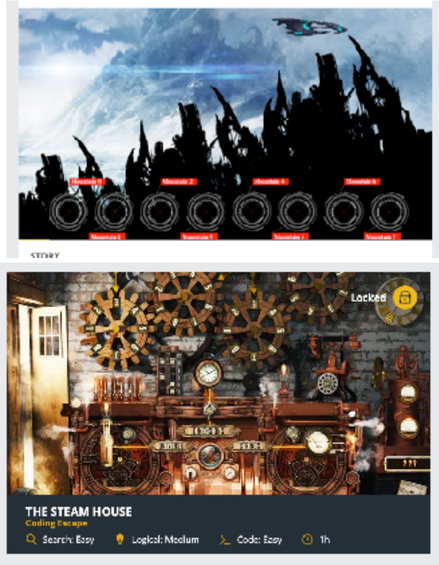
- you are the only one responsible for your path/future so think about it NOW!
- you do not need network nor luck -> you need skills/knowledge (SQL, Python, …)!
- plan your project, measure impact and always keep the bigger view in mind
- propose change and challenge -> do it by example
- find out what do you like to do and go this path!
- start small and build on top of that -> do not over-engineer (not right away)
- learn Git and Python and SQL - by your own projects ! (this gives flexibility for any carer path in tech)
- remember PhD is a real experience (you learn lots of things) -> business needs to accept it!
- when doing something be able to estimate what/when it is a success and when failure
- please don't prepare you CV in TeX :)
- always think bigger -> do not depend on technology, network, luck
- 1 job offer = 1 customised CV
- train the recruitment process !

THANK YOU!

Recommendations:

CodinGame

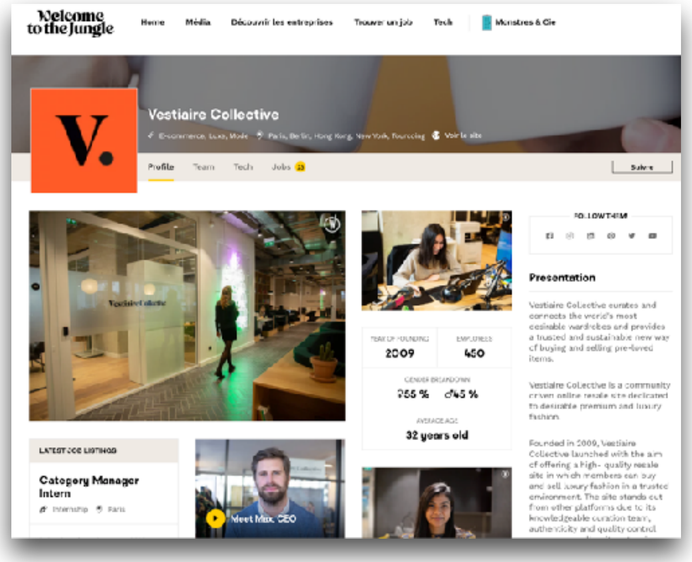
www.codingame.com



- [ ] Bash (3)
- [ ] C# (83)
- [ ] C++ (43)
- [ ] CSS (12)
- [ ] Elixir (3)
- [ ] F# (1)
- [ ] Go (1)
- [ ] Groovy (1)
- [ ] Haskell (4)
- [ ] HTML (7)
- [ ] Java (91)
- [ ] Javascript (93)
- [ ] Kotlin (18)
- [ ] .Net (16)
- [ ] PHP (16)
- [ ] Python (86)
- [ ] React.js (9)
- [ ] Ruby (5)
- [ ] Rust (5)
- [ ] Scala (4)
- [ ] spring (10)
- [ ] Swift (4)
- [ ] Other (483)

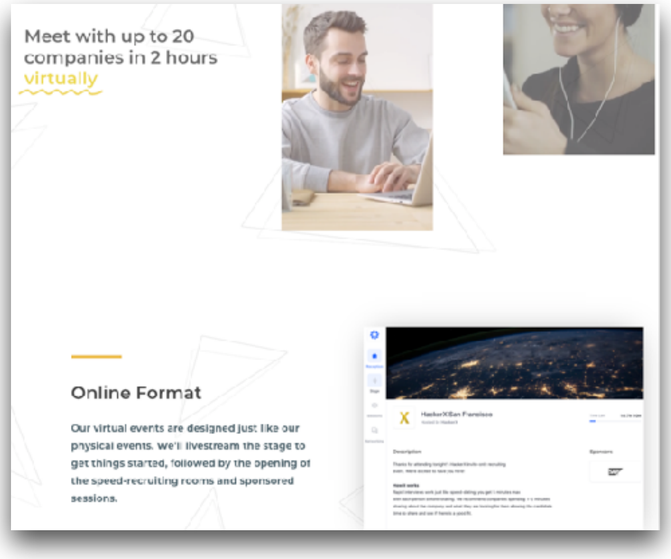
Welcome to the Jungle

www.welcometothejungle.com



HACKERX

https://hackerx.org/



THANK YOU!