AI & Robotics

Machine Learning / AI PE

Visite São Paulo sem o Primeiro Comando da Capital!

Intro

- Visit sunny São Paulo!
- Ch[illegible]c city

Tasks

- Data processing
- Data visualization
- Prediction and exploration
- Website and recommendation

Datasets

	Weather	Crime
Daterange	2000-2016	2007-2016
# Records	~10 million	~15 million
Size (MB)	~2000MB	~8000MB
Link	Weather dataset	Crime dataset

Tasks: Data processing

	9779163	9779164	9779165	9779166	9779167
wsid	423	423	423	423	423
wsnm	BARUERI	BARUERI	BARUERI	BARUERI	BARUERI
elvt	777	777	777	777	777
lat	-23.5239	-23.5239	-23.5239	-23.5239	-23.5239
lon	-46.8695	-46.8695	-46.8695	-46.8695	-46.8695
inme	A755	A755	A755	A755	A755
city	Barueri	Barueri	Barueri	Barueri	Barueri
prov	SP	SP	SP	SP	SP
mdct	2016-09-30 19:00:00	2016-09-30 20:00:00	2016-09-30 21:00:00	2016-09-30 22:00:00	2016-09-30 23:00:00
date	2016-09-30 00:00:00	2016-09-30 00:00:00	2016-09-30 00:00:00	2016-09-30 00:00:00	2016-09-30 00:00:00
yr	2016	2016	2016	2016	2016
mo	9	9	9	9	9
da	30	30	30	30	30
hr	19	20	21	22	23
prcp	NaN	NaN	NaN	NaN	NaN
stp	927.6	928.1	928.7	929.6	930.5
smax	927.7	928.2	928.7	929.6	930.5
smin	927.4	927.5	928.1	928.7	929.5
gbrd	599.96	243.923	52.334	0	NaN
temp	16.3	15.3	15	14.6	14.3

1st task: get relevant insights out of the datasets

Weather

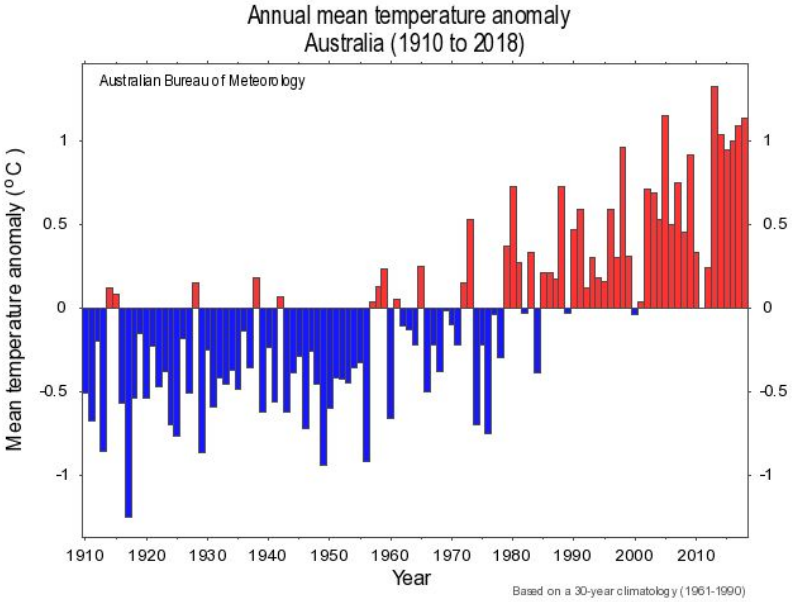
- Temperature data
  - Monthly/yearly average
  - Min temperature, max temperature
- Draught periods
- Climate change impact
- …

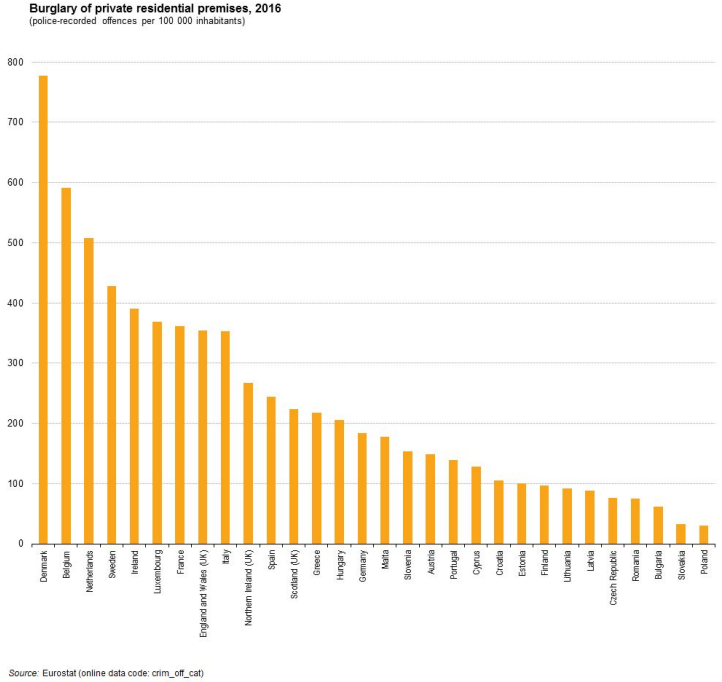
Crime

- What regions of the city are the worst for violent crimes?
- Has there been an increase or decrease in violent crime throughout the years?
- Which population groups are more likely to be perpetrators?
- ...

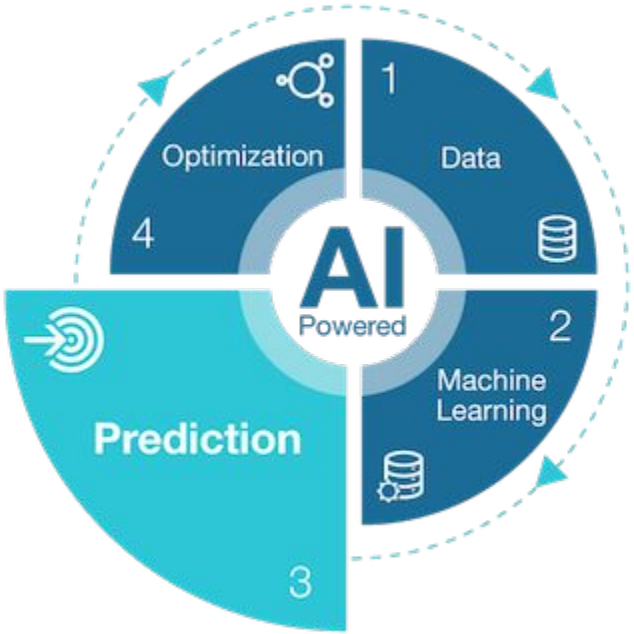
Tasks: Data visualization

2nd task: meaningful data visualization





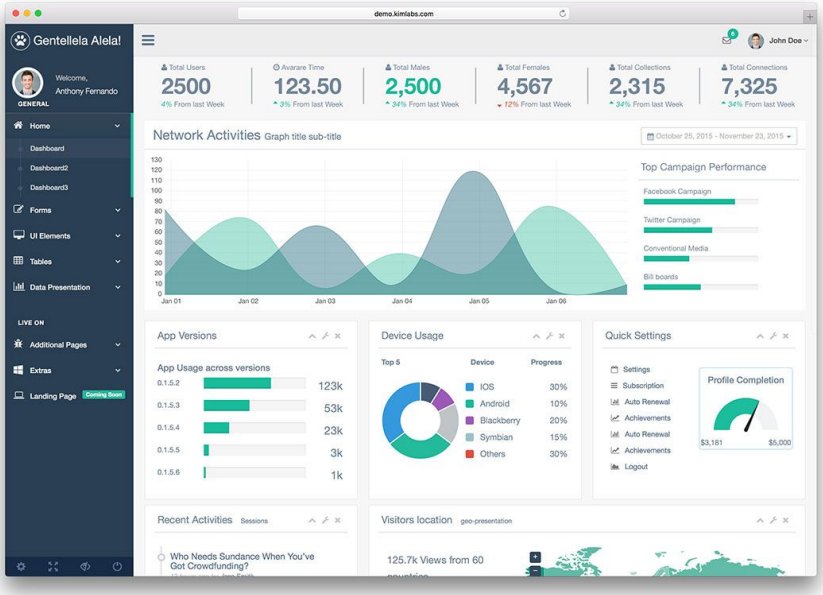
Tasks: Prediction and exploration



3rd task:

- Generate predictions for 2017
  - For weather data
  - For crime data
- Use techniques from the course
  - Random Forests
  - Neural Networks
  - Clustering
  - Or try your own solutions! (EXTRA!)
- Optimize and compare your models
- Find correlations in the data

Tasks: Website and recommendation



4th task:

- Construct a REST API and a front-end displaying the different data visualizations you have come up with
- Recommend a tourist wishing to visit São Paulo
  - When is it optimal to visit this city?
  - Which parts of the city to avoid?
  - ...

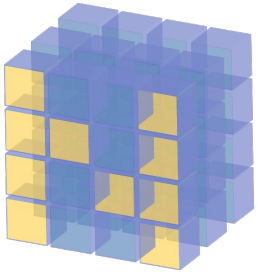
  => The tourist has to supply basic information like:

  - Gender
  - Age
  - Race (Yes we know, but it's statistically relevant.)
  - ...

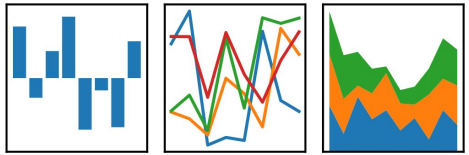
Keep it simple!

Use templates!

Tools



NumPy



Evaluation

%	Periode	Type
40	June	Group project (individual assessment)
60	June	Written Exam (closed book)
40	Aug/Sep	Individual assignment
60	Aug/Sep	Written Exam (closed book)

Evaluation

- Data analysis / ML code
- Working REST API and front end
- Presentation + demo
- Written report (~5 pages)
  - → `README.md`
  - → Short summary
  - → Retrospective++ of the project
    - → Mad, Sad, Glad?
    - → Pitfalls?
  - → Personal reflection of each member

Deadline

- Every week there is a block of time during class allocated to the PE
- Deadline: Last commit: 28/05/2019 at midnight! Commit often!
  (Add a subdir to your Research Project repository called AIPE)
- Presentation: 29/05/2019