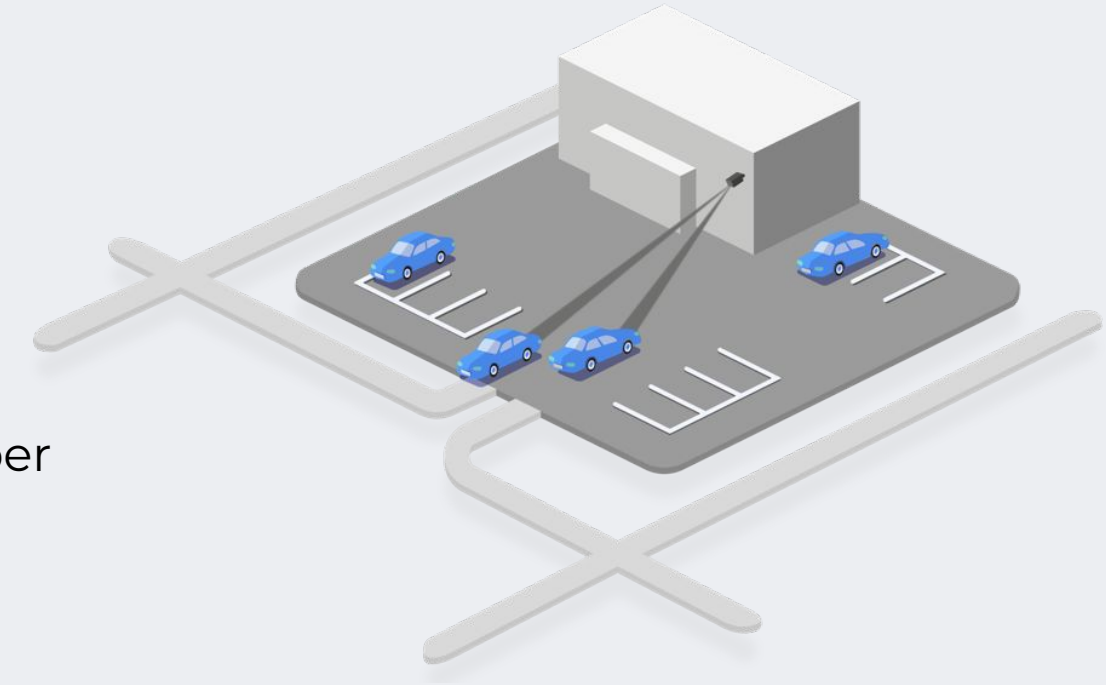
Case Study: Data Analyst

Introduction: Welcome to the Parkdepot Case Study

Focus: Data analysis & visualization
Tools: Python or R (whichever you feel more comfortable with)
Presentation: Please choose any form of presentation, where we can see both the code and structured output / visualizations (e.g. Jupyter Notebook, R Markdown, Markdown file in your IDE)
Presentation time: 20 - 30 minutes
Preparation time: Please don’t invest more than 5 hours of your time in this
Documentation: Please send us a folder containing your script, a short readme and everything else we need in order to run your script
Language: English

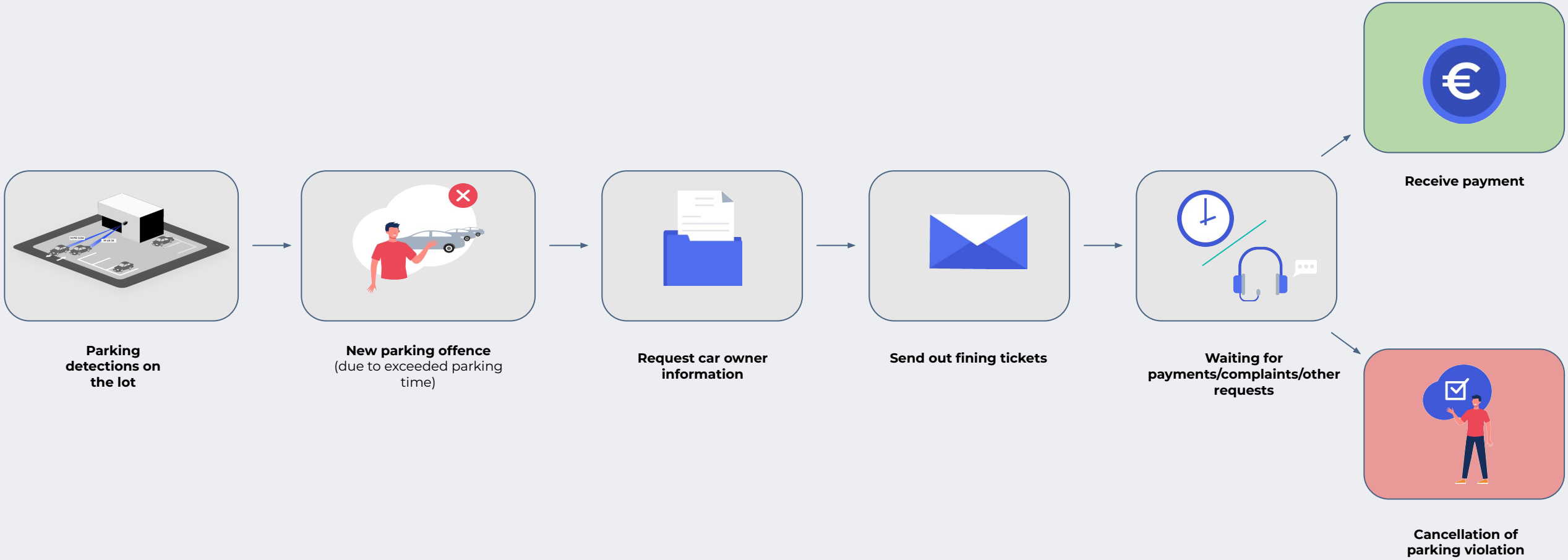
Introduction: Enforcement at Parkdepot

- Installation of the parking system at the entrances and exits of the parking areas
- Automated registration of vehicles and number plates on entry and exit
- Parkdepot records third-party parkers, analyses their parking durations and collects fines from parking offenders who exceed the allowed parking duration



Introduction: The scanner system

Introduction: Offender persecution process



Challenges

Challenge I: Information about the data

- Please assume that dataset 1 was exported on **June 7, 2016** and dataset 2 on **July 4, 2016**

- The “life-cycle” of a offence case is tracked by its status (see slide 3).
  The ideal status-chain would be the following:
  1. **new** (new violation where we need to figure out the owner)
  2. **waiting_for_owner_data** (requested owner information and waiting for answer)
  3. **ready_to_send_letter** (letter for violation can be sent out)
  4. **letter_sent** (first letter was sent to parking-violator in which we claim a parking-fine)
  5. **paid** (the parking violator paid the fine after he/she received the letter)

- If an offender does not pay after a specific amount of days, the person will receive additional letters

- Canceled cases always have a cancellation reason and date. That way each cancellation can be categorized

- The Extra-charges column combines potential administration and/or warning fees

Challenge I: Offender persecution at Parkdepot

Given are two datasets with mocked data of detected parking-offenders of an entire month each. Between dataset 1 & 2 lay 6 months.

Questions:

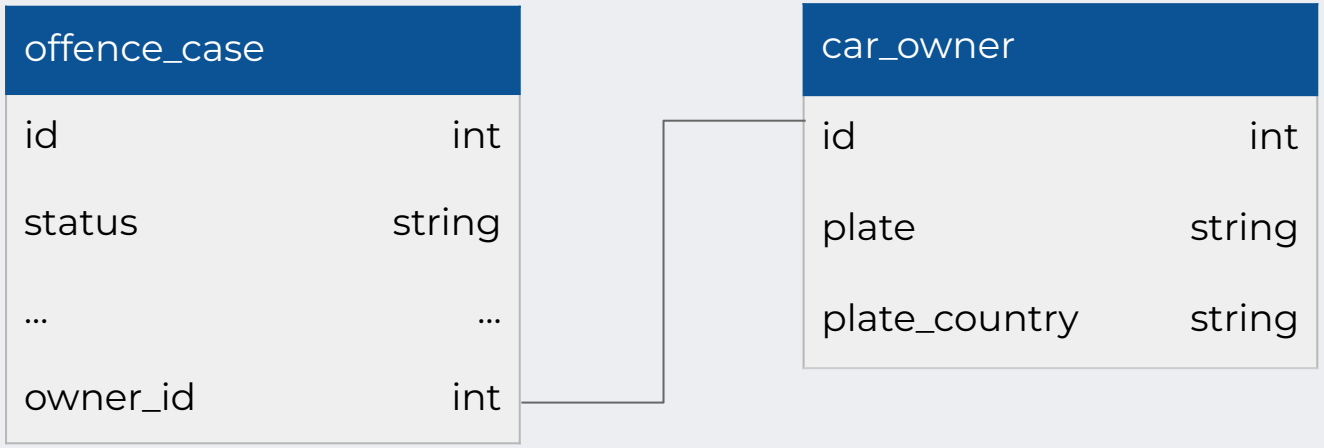
- How would you define the payment rate of all offences?
- Which additional metrics could be interesting?
  Please create 2 graphs of your choice that represent what you believe to be key performance indicators of the business (e.g. customer, product, revenue (...) driven)
- What can you find out about offence specific statistics and general behavior of parking-offenders?
- What has changed during the time between the first and second data set?
- Where do you see the most urgent need for improvement in our system?
- What additional data would be useful to have?

Underline your analysis and argumentation with visualizations of your choosing or convincing numbers.

Challenge II: Data querying

Based on the data model below, write a SQL query to select all paid offences of international (non German) car owners.

Bonus: How would the query look like if you would only like to select paid cases of offenders who have caused at least two paid offences?




Parkdepot
Have fun and happy coding!