Car Sharing Analytics

Danilo Giordano - danilo.giordano@polito.it

Indaco Biazzo – indaco.biazzo@polito.it

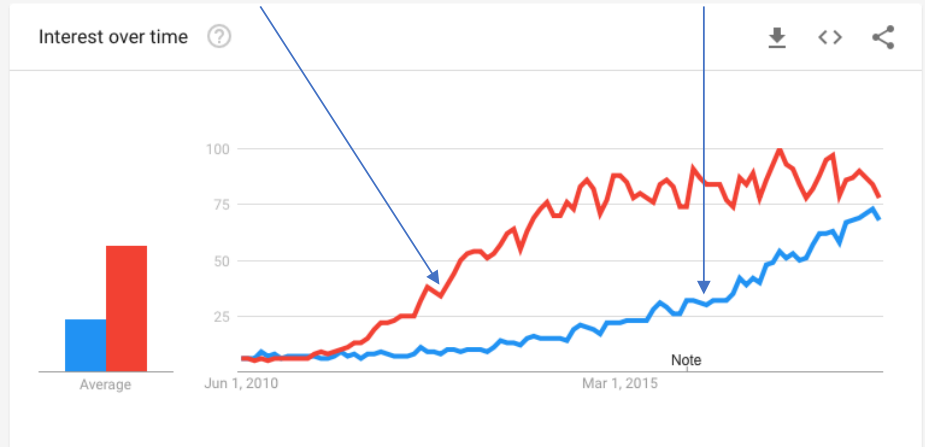
SmartData@PoliTo

Inter-departement center

https://smartdata.polito.it

Big Data & Data Science

"Extracting **meaning** from very large quantities of **data**"

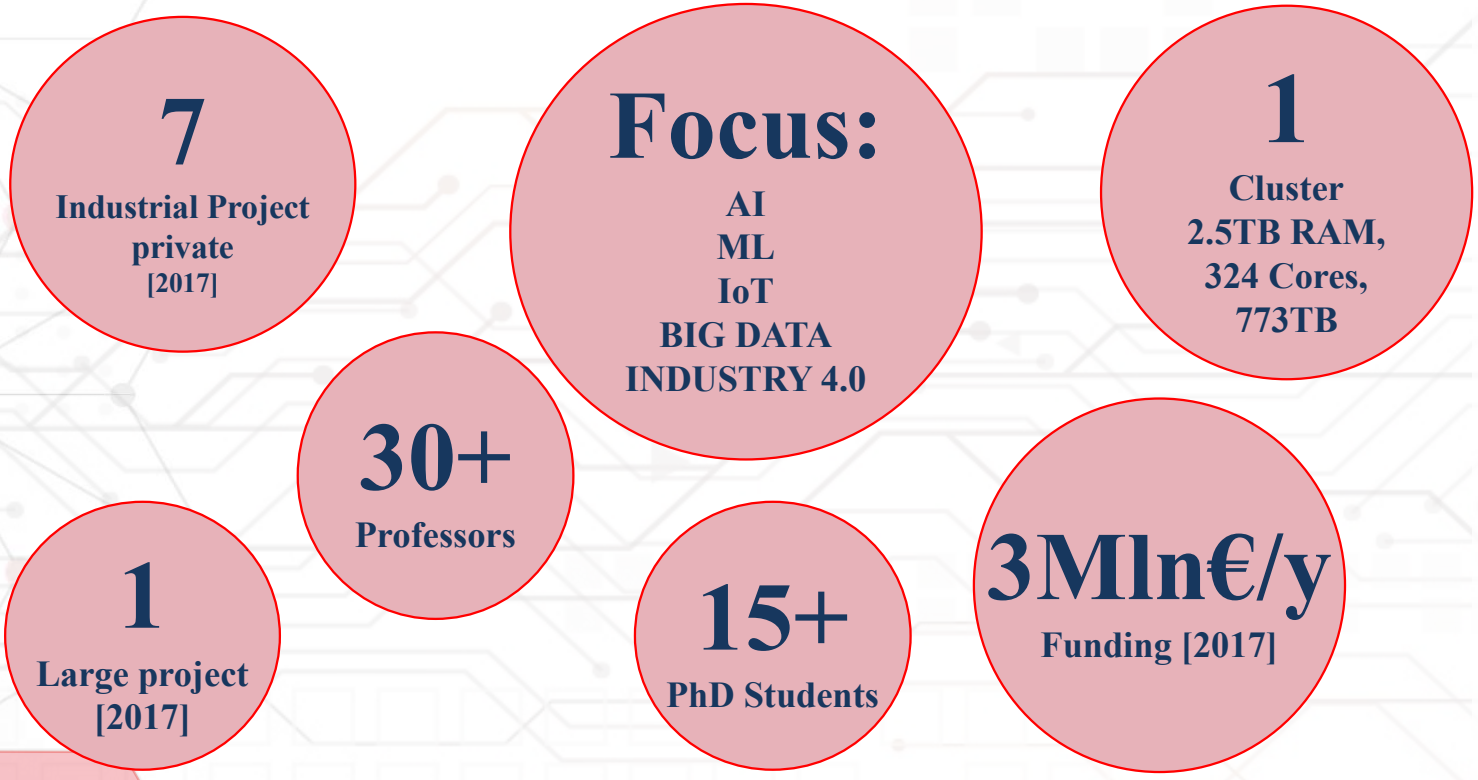


GOAL of SmartData@PoliTO

- Born on the experiences and competencies
  - EE – Internet monitoring
  - CS – Data mining & machine learning
  - MAT – Statistics and modelling
  - Phy – Modelling and inverse problems
  - MNG – Economics and business impact
  - …
- And on the BigData@PoliTO centre
  - Fully dedicated Cluster hardware
- **Offer access and support** for these technologies
- **Technology transfer:** toward industries

SmartData@PoliTO in one slide

Data updated May 2018



Information

Complete Name
SmartData@PoliTO

Parent organization
Politecnico di Torino

Sponsors
Both Private and public

Location
Torino

Born in
2017

Website
smartdata.polito.it

Coordinator
Marco Mellia

Address
Corso Duca degli Abruzzi 24 - 10129 Torino

Contacts
contact.smartdata@polito.it

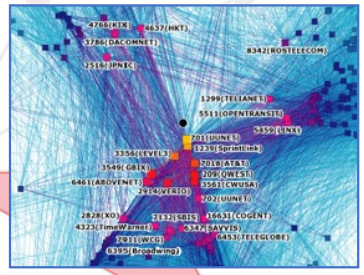
Competences

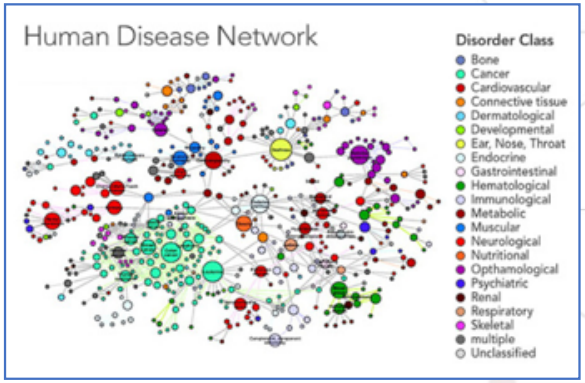
- **Algorithms** and **methodologies** for data analysis
  - Big data processing, data mining, supervised and unsupervised machine learning, deep learning, rule mining, classification, prediction, anomaly detection, clustering, …
- **Methodologies** for **data modeling**
  - Statistics, computational topology, geometry with application to data analysis, graph modeling and information mining from graphs, …
- Study the **transformations** in **industrial structures**
  - Value chains, business models, managerial practices, …

Generation → Acquisition → Storage → Analysis

Applications and domains

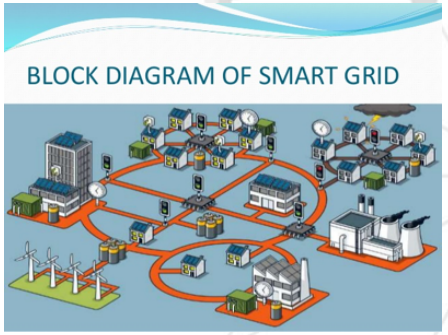
- Predictive maintenance for complex systems
- Anomaly detection with applications to time series analysis
- Cybersecurity
- Mobility in smart cities
- Data diffusion in complex systems

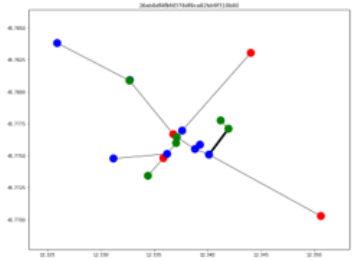


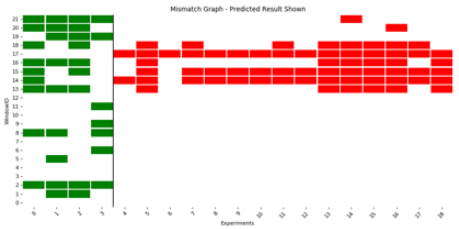


Collaboration with Private Companies

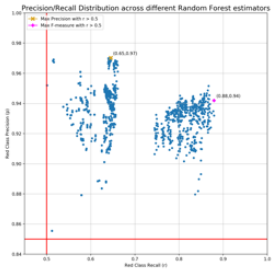
- **Predictive Maintenance**
  - Analysis of data from **electric grid** (SCADA) to predict **failures**/malfunctioning of vital elements







- **Advanced Diagnostic/Predictive Maintenance**
  - Analysis of data from **powertrain systems** to predict **failures**/malfunctioning



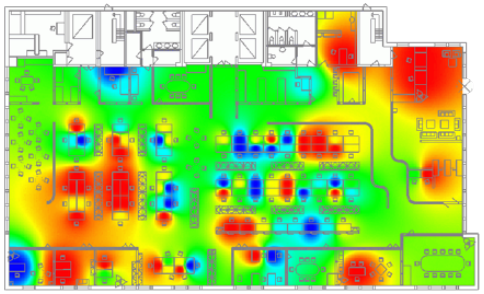
Collaboration with Private Companies

POLITECNICO DI TORINO

SmartData

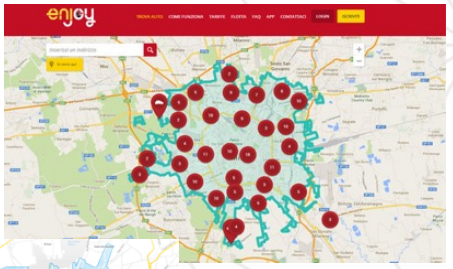
EDISON

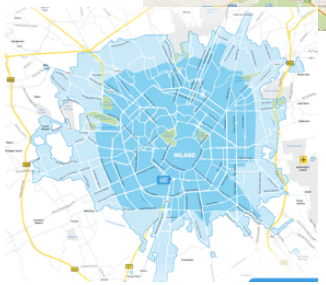
- **Smart City**
  - Harvesting of **open data** for the characterization of energy consumption of residential and commercial **buildings**



SIEBEL ENERGY INSTITUTE

- **prediction**
  - Study of **traffic patterns** from car sharing systems
  - Design of **electric car sharing system**





Can we use car sharing data to…

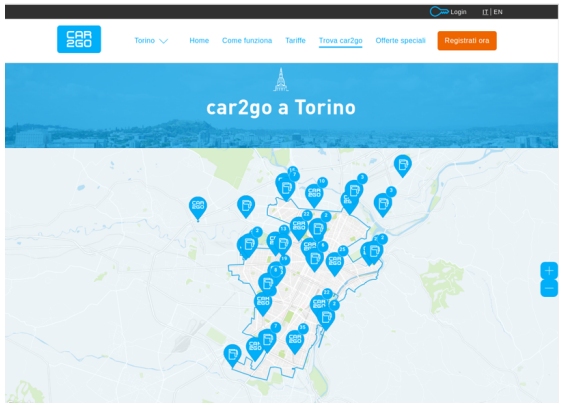
- … Understand user behavior?
- … Improve system design?
- … study novel solutions?

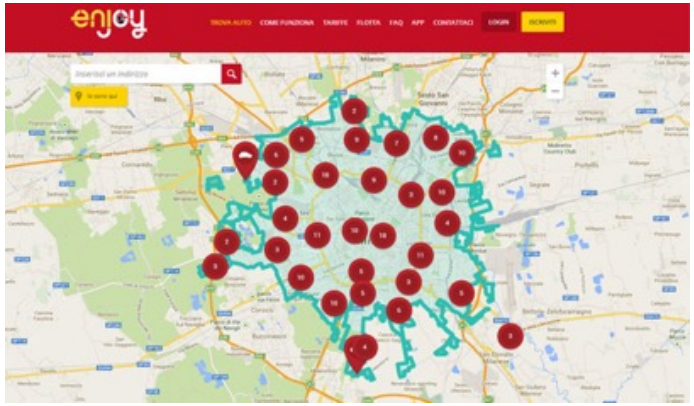
- For example:
  - In a future electric car sharing system, where the user is not force to plug the car all the time, **how shall I place the charging station?**

Car Sharing

- Free Floating Car Sharing (FFCS) system
  - Users can pick and drop a car wherever in a geo-fence area
  - Car2go and enjoy

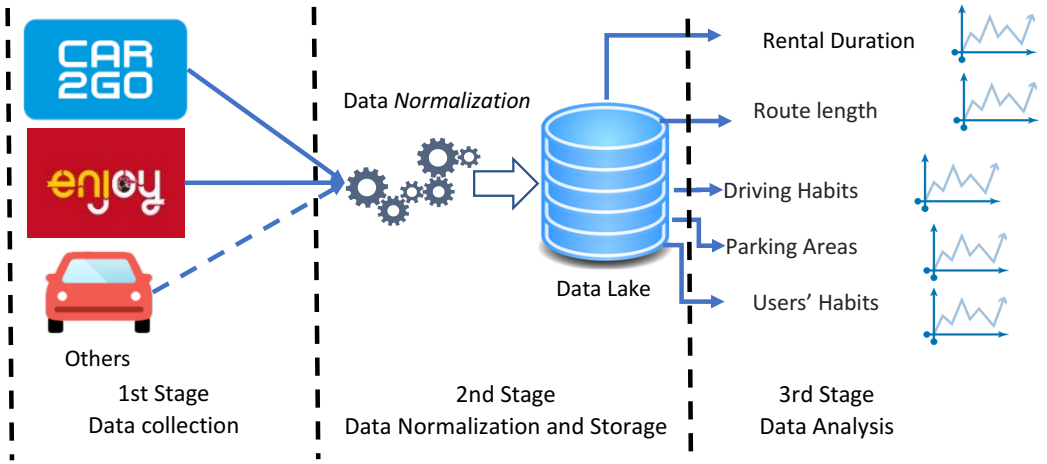
… How do we get car sharing data?





Data Collation and Processing

- 3 stage framework
  1. **Data Collection**: **Collect** data from **different FFCS** providers
  2. **Data Normalization and Storage**: **Process** the collected data to **normalize** it in a **common format** and save it in a **Data Lake**
  3. **Data Analysis**: **Analyze** the data to extract **meaningful** information



1st Stage: Data Collection

- 2 Crawlers fetch data from 2 FFCS providers
  - Every minute we get a new **snapshot** of available cars
- **car2go**
  - Public API available until January 2018[1]
  - Based on a shared key in the URL
  - Download list of available cars in a specific city as a JSON document
    - The city is directly specified in the URL
- **Enjoy**
  - NO API available
  - Use cookies to create a session and download the data
  - Download list of available cars in a specific city as a JSON document
    - The city is specified in the cookies
- JSON document
  - List of available car plus for each car several information

[1]: https://github.com/car2go/openAPI

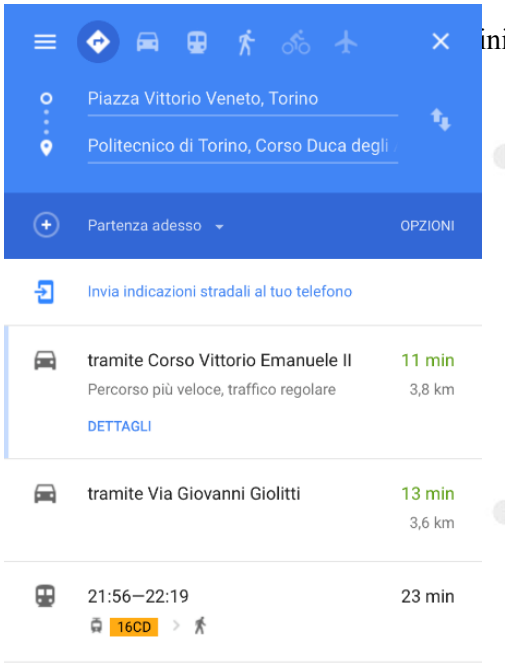
Example of a JSON document


```
{
    “Plate" : "292/EZ727ZW",
    "Vin" : "WME4513341K725150",
    "Coordinates" : [7.68461, 45.05523],
    "Fuel" : 100,
    "Address" : "Corso Massimo d'Azeglio 26, 10125 Torino",
    …
    }
```

2nd Stage: Data Normalization and Storage

- Process the crawler data to extract **bookings**
  - **The bookings** keep information about the *bookings* of the cars i.e., when a car was (likely) used by a customer
  - Save Car2go and Enjoy data in a **common** format
  - Many information such as: the origin coordinates, the destinations initial time, the final time, the car plate, etc…

- Integrate the data with other sources
  - Use google map API to integrate the data
    - Given a booking from [x0,y0] to [x1,y1]
    - When a booking ends, query google maps API for data
      - Query limitations due to the number of (free) queries
  - Which was the possible path the driver followed?
  - Which should be the minimum time to complete journey?
  - How long would it take to walk instead of driving?
  - How long would it take by using public transport means?



Caveats

In reality, things are more complicated

- **Problem:** A car may disappear, and reappear in the same place
  - GPS fix position error
  - Booking that has been cancelled
  - System issues?
- **Solution:** Keep bookings longer than **700 m**
- **Problem**: Car may disappear for (very) long time
  - Car brought to maintenance
- **Solution:** Keep bookings lasting less than **1 h**
- **Problem**: A lot of car may disappear at the same time
  - Maintenance?
  - System issues?
- **Solution**: If more than 30% car disappear do not modify any temporarily booking or parking
- **Problem**: A car may reappear in a different city
  - Relocation?
  - System issues?
- **Solution:** Keep bookings shorter than **45 km**

- Google maps data may be missing
- …

Always double check what you get!

Filtering wrap-up

- Filter ***bookings*** to obtain ***rentals***
  - **Distance** > 700 m and **Distance** < 45 km
  - **Duration** < 1 h

- **Important:** Duration = Reservation Time + Driving Time + Parking Time
  - **Reservation Time:** time from the moment the car was booked by the user up to the start driving moment (**unknown**)
  - **Driving Time:** actual driving duration (**estimated** with google map API)
  - **Parking Time:** time spent by the user looking for a new parking spot (**unknown**)

Which information describe the rental?

How many rentals do we have?

Dataset

- Data is structured in tabular format coded in csv
- Each field is divided by ;
- Each field is enclosed with “”
- Each rental is described by 25 columns

E.g.,

"865775";"enjoy";"Torino";"EZ117GW";"-";"Fiat 500";"1514761298";"2018-01-01 00:01:38";"34";"45.107143";"7.7032084";"45.10154";"7.694324";"1514762999";"2018-01-01 00:29:59";"31";"Corso Giulio Cesare, 299, 10155 Torino TO";"Corso Vercelli, 193, 10155 Torino TO";"940";"1203";"523";"1635";"1514764185";"251";"1220"

Rental Information

POLITECNICO DI TORINO

SmartData

Name	Meaning	Example
id	Unique rental identifier [progressive number]	1
vendor	Car sharing provider [car2go, enjoy]	Enjoy
city	City of the rental [Torino, Milano, Roma]	Torino
plate	Car unique identifier	EZ117GW
vin	Car unique identifier [**only for car2go**]	-
car_name	Vehicle manufacturer and model [**only for enjoy**]	Fiat 500
init_time	Timestamp when the rental started	1514761298
init_date	Date of when the rental started [from timestamp]	2018-01-01 00:01:38
init_fuel	Fuel level at the beginning of the rental [in percentage]	34

Rental Information 2/3

Name	Meaning	Example
origin_lat	Latitude of where the rental started	45.107143
origin_lon	Longitude of where the rental started	7.7032084
dest_lat	Latitude of where the rental ended	45.10154
dest_lon	Longitude of where the rental ended	7.694324
final_time	Timestamp when the rental ended	1514762999
final_date	Date of when the rental ended [from timestamp]	2018-01-01 00:29:59
final_fuel	Fuel level at the end of the rental [in percentage]	31
init_address	Address of where the rental started	Corso Giulio Cesare, 299, 10155 Torino TO
final_address	Address of where the rental ended	Corso Vercelli, 193, 10155 Torino TO

Rental Information 3/3

Data integrated with Google Maps API

Name	Meaning	Example
walking_duration*	Duration of the trip from the origin to the destination if the user walked [seconds]	940
walking_distance*	Distance of the trip from the origin to the destination if the user walked [meters]	1203
pt_duration*	Duration of the trip from the origin to the destination if the user used public transport means [seconds]	523
pt_distance*	Distance of the trip from the origin to the destination if the user used public transport means [meters]	1635
pt_arrival*	**Timestamp** of arrival if the user used the public transport means. **Required** to correctly estimate duration in **night** trips	1514764185
driving_duration*	Duration of the trip from the origin to the destination if the user drove [seconds]	251
driving_distance*	Distance of the trip from the origin to the destination if the user drove [meters]	1220

*if no information is available the field is set to **-1**

Dataset

Rentals per Month

Period	Torino*		Milano*		Roma	
Provider	**car2go**	**Enjoy**	**car2go**	**Enjoy**	**car2go**	**Enjoy**
June 2017	50641	60047	125238	159952	62384	99985
June 2017	50136	46436	125712	115933	68868	57663
January 2018	64223	48549	157578	152013	74506	97004
February 2018	0	37818	0	111118	0	77878

*City for which we have Google Maps information

Fleet size

Provider	Torino	Milano	Roma
Car2go	410/470	780/820	590/610
Enjoy	260/500	920/1480	600/1310

January Fleet size

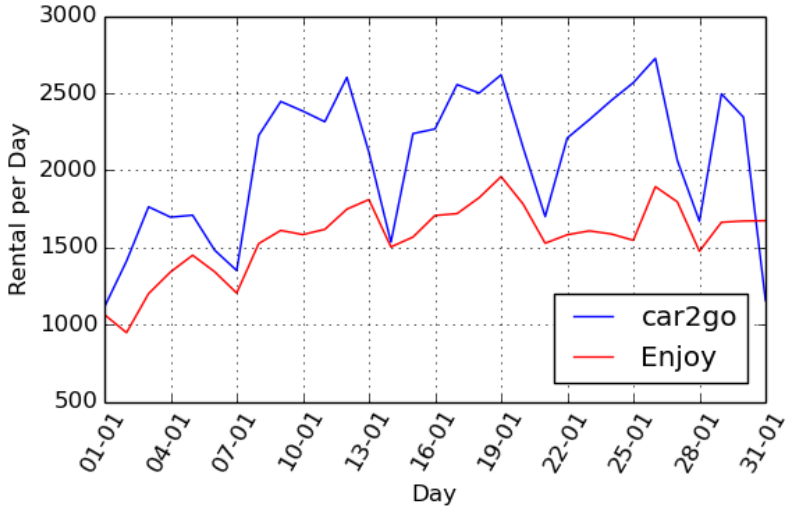
Provider	Torino
Car2go	413
Enjoy	279

A few results

Rental per day

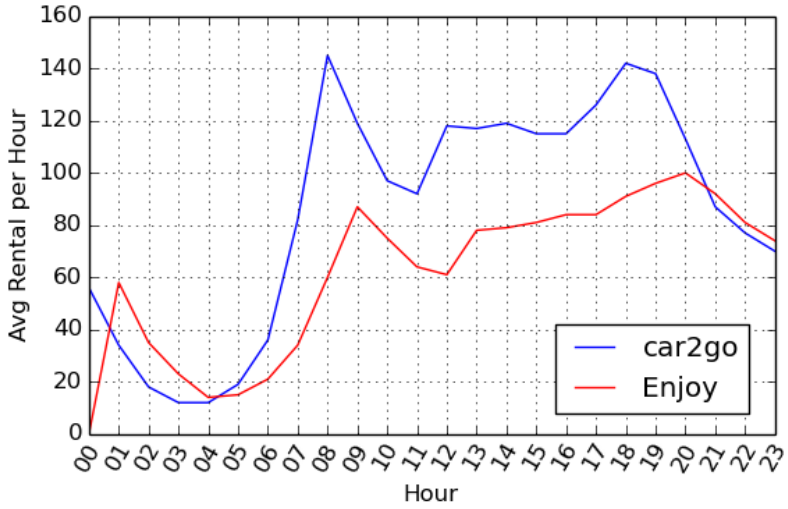
- January 2018
  - Car2go: 413 cars
  - Enjoy: 279 cars

- car2go utilization highly change during the month while Enjoy utilization is more stable
- Ca2go utilization drops during the weekend days and festivity
- **Insight**: people may prefer
  - 2 sits car during the weekdays for business reasons
    - E.g., faster find a parking spot to go to work
  - 4 sits car otherwise



Rental per hour

- Each bin of 1 hour is the average number of rentals over the month
- Despite **Enjoy** fleet is smaller, after 8 pm it is more used then **car2go**
- **Insight**
  - Car2go is used more during commuting hours
  - During evening people prefer to share the ride taking a 4 seats model

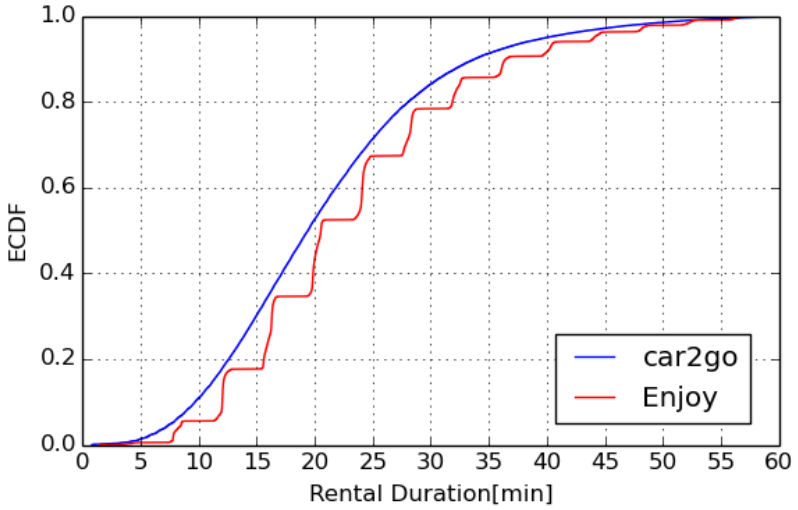


ECDF Rental Duration

- **Recall:** Rental Duration = Reservation Time + Driving Time + Parking Time

- Enjoy shows a discrete pattern (steps) due to asynchronous updates
  - Extra delay at booking end…
- Car2go and Enjoy have similar distributions
- 90% of rentals last less than 40 min
  - **This time includes reservation and parking time**

- **Insight**
  - Most of the customers use FFCS for short trips in duration

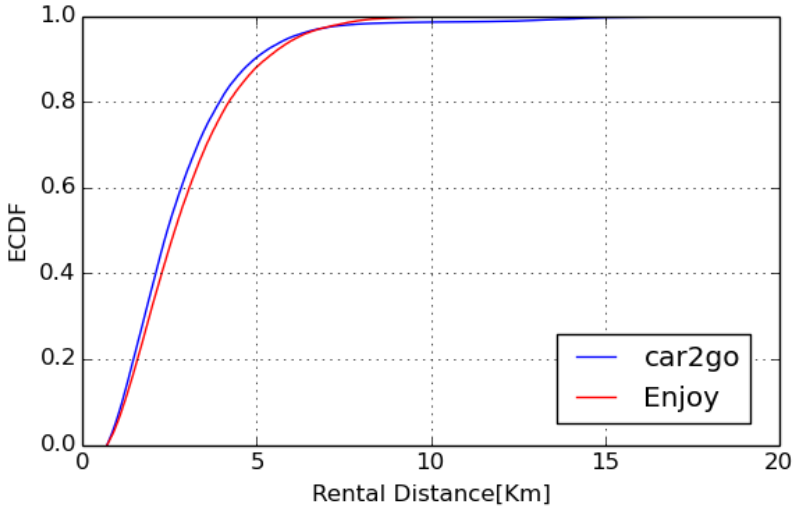


ECDF Rental Distance

- The rental distance is computed as the Euclidian distance between origin and destination coordinates

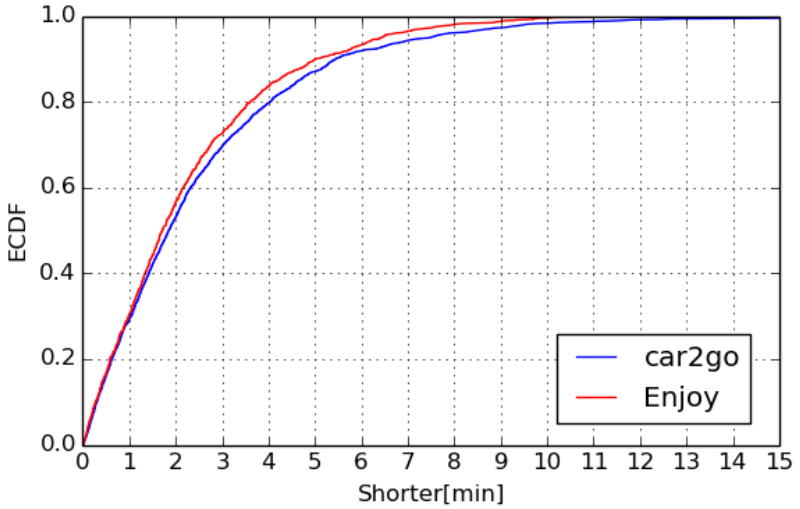
- Driving distance is shorted than 5 km 90% of the time
- As with Car2go it is possible to reach the airport the maximum rental distance is higher

- **Insight**
  - Most of the customers use FFCS for short trips in distance as well



ECDF Rental Time shorter Driving Time

- When the Google maps data are available compute the difference between **driving duration** and the **rental duration**
- **Recall**: **Rental Duration** = Reservation Time + Driving Time + Parking Time
- Trip with **driving duration** shorted than rental duration are:
  - 12.0% for car2go
  - 6.5% for Enjoy
- 50% of the time the trip is shorter it is 2 minutes shorter
- **Insight**
  - Car2go drivers seems to drive faster than Enjoy drivers

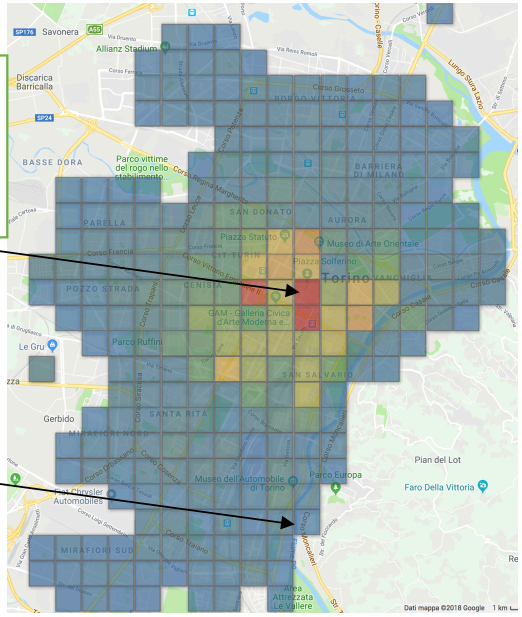


Users' parking habits

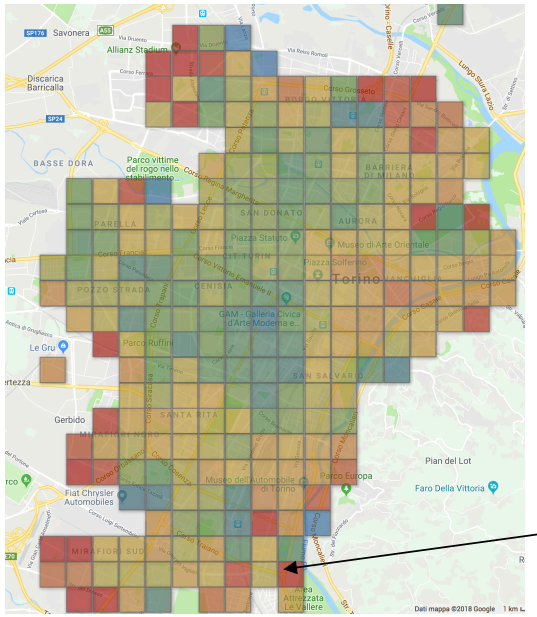
- We divide Turin in a grid with square zones of 500 m length
- To each zone we compute: **Number of parking** and **Average parking** Time

Porta Nuova
(main railway station)
47.15 rentals/day

Periphery,
0.017rentals/day



Number of parking



Average parking time

1 parking lasted 31 h

Challenges

Additional material

- You can find this presentation, the challenges presentation, and the Python notebook description @
  - https://github.com/CityChrone/BigDive/tree/master/presentation

- All the material is in the GitHub repository
  - https://github.com/CityChrone/BigDive