



Centre Tecnològic de Catalunya

| Data Science ❤️ Agility

a tale of three small and agile projects



Pierdomenico Fiadino, PhD
Big Data Analytics Unit at EURECAT

Innovant amb les empreses



The Uber case



Quick facts

- Founded in 2009
- Operations in >600 cities
- Employees >12000
- Revenues (in 2016)
- Riders (in 2016)



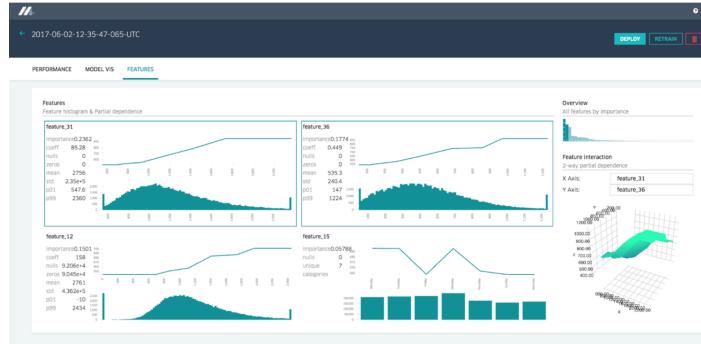
appetite comes eating...

UberEATS

- Mobile meal ordering service
- Launched in San Francisco ('14)
- >30 minutes delivery
- 5M downloads (Play Store)



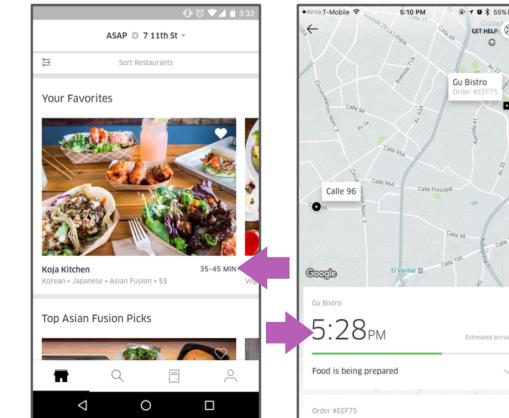
Uber: big A.I. investments



Estimation of preparation time for UberEats

Michelangelo

- Uber's ML platform
- State-of-the-art technologies
- HDFS, Spark, TensorFlow, ...
- AI Labs (Seattle, 3000+ engineers)



source: eng.uber.com/michelangelo

What about smaller/shorter projects?
say, 2-3 months, small team (1-5)



***“Judge me by my size,
do you?”***

- Master Yoda

First story: **Alphanet**

building BI-capable and extensible Big Data platforms

Quick and lean

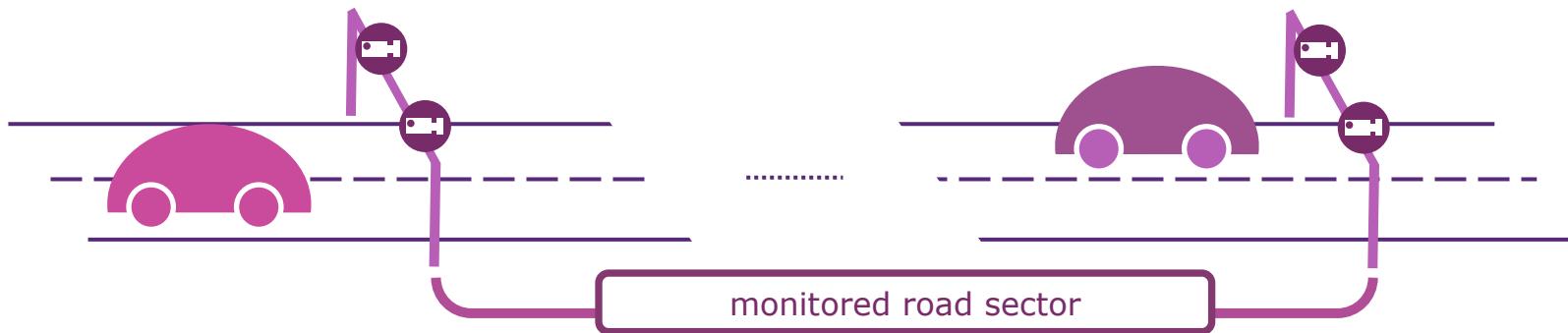
Road cameras: security and traffic



About Alphanet

- Catalan startup
- Road monitoring & security
- >200 monitored road segments

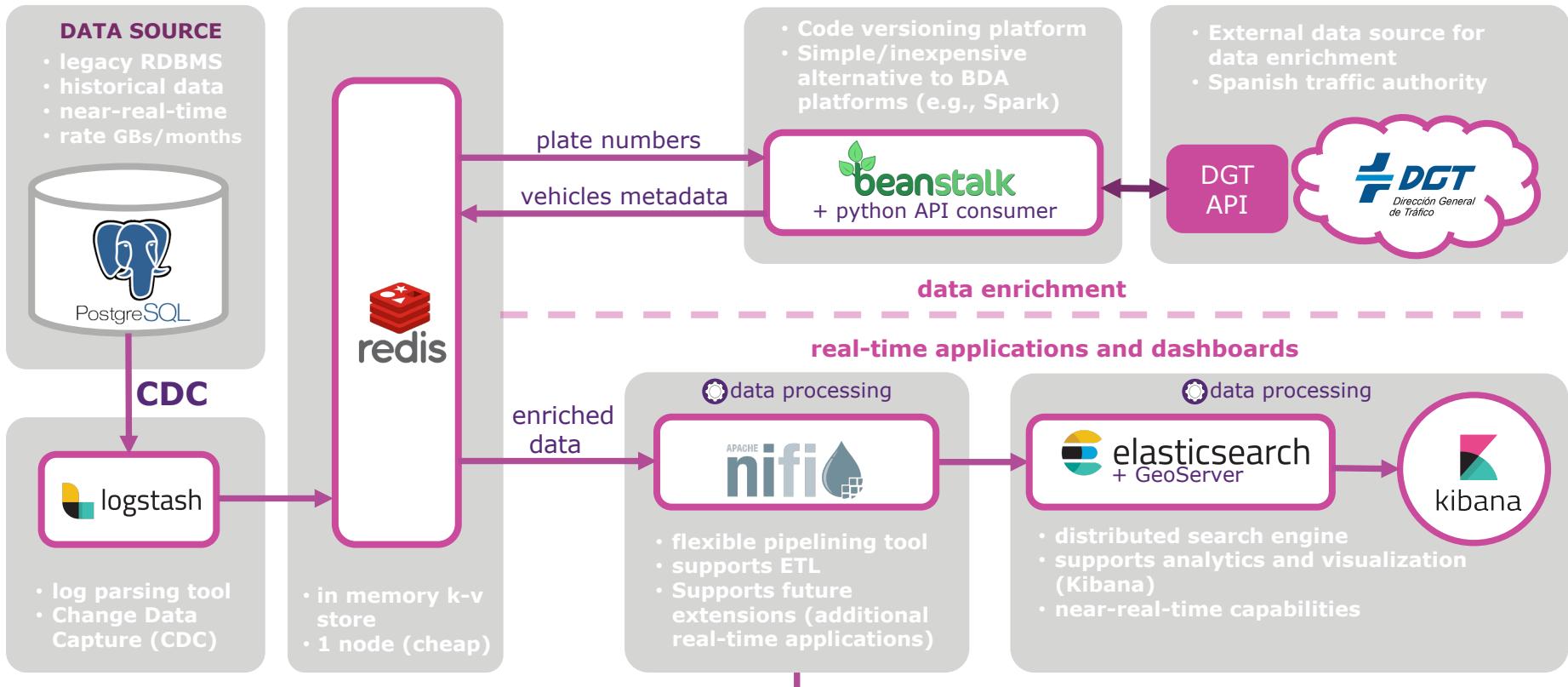
project goal: **deploy a BI platforms to allow visualization of traffic statistics**



tens of cameras, tens of monitored sectors

Goal: MVP platform

Alphanet: data pipe-line



Alphanet: conclusions



Data aggregations, support for traffic analytics

Supports extensions



Dashboard

Kibana-based



Inexpensive cloud infrastructure

MVP at the end of the project



Project duration: 3 months

building infrastructure and algorithms



600€/month



600€/month



400€/month



250€/month

Second story: **012.cat**

Citizen information office

Data characterization
and informed actions

012: Citizen Information Service

012
gencat.cat

Telephone Information Service for Citizens

- Managed by Generalitat de Catalunya
- Paid service (time-based billing)
- Answers to questions of different topics
 - health, tax-related, wellness, welfare, traffic, ...



Project goal: **study call routing process**



012: dive into historical data

dataset

- time span
- number of records
- attributes
- size

1 year
800K
>10GB

Call time

(time of the day, day of the week, etc.)

Call length

(waiting/talking time, etc.)

User info (when available)

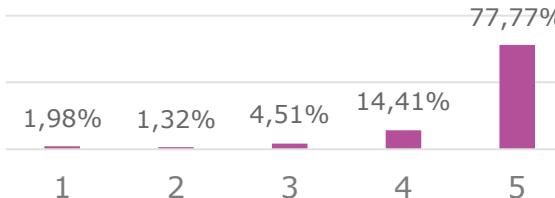
(geographic position, .)

Other call metadata

(topics, routing, agents, ratings, etc.)



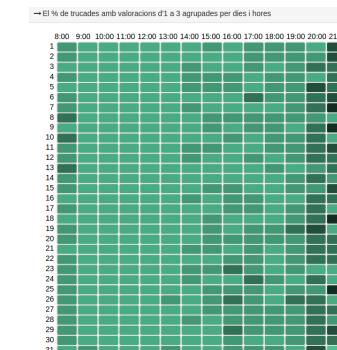
statistical characteristics



how ratings are distributed

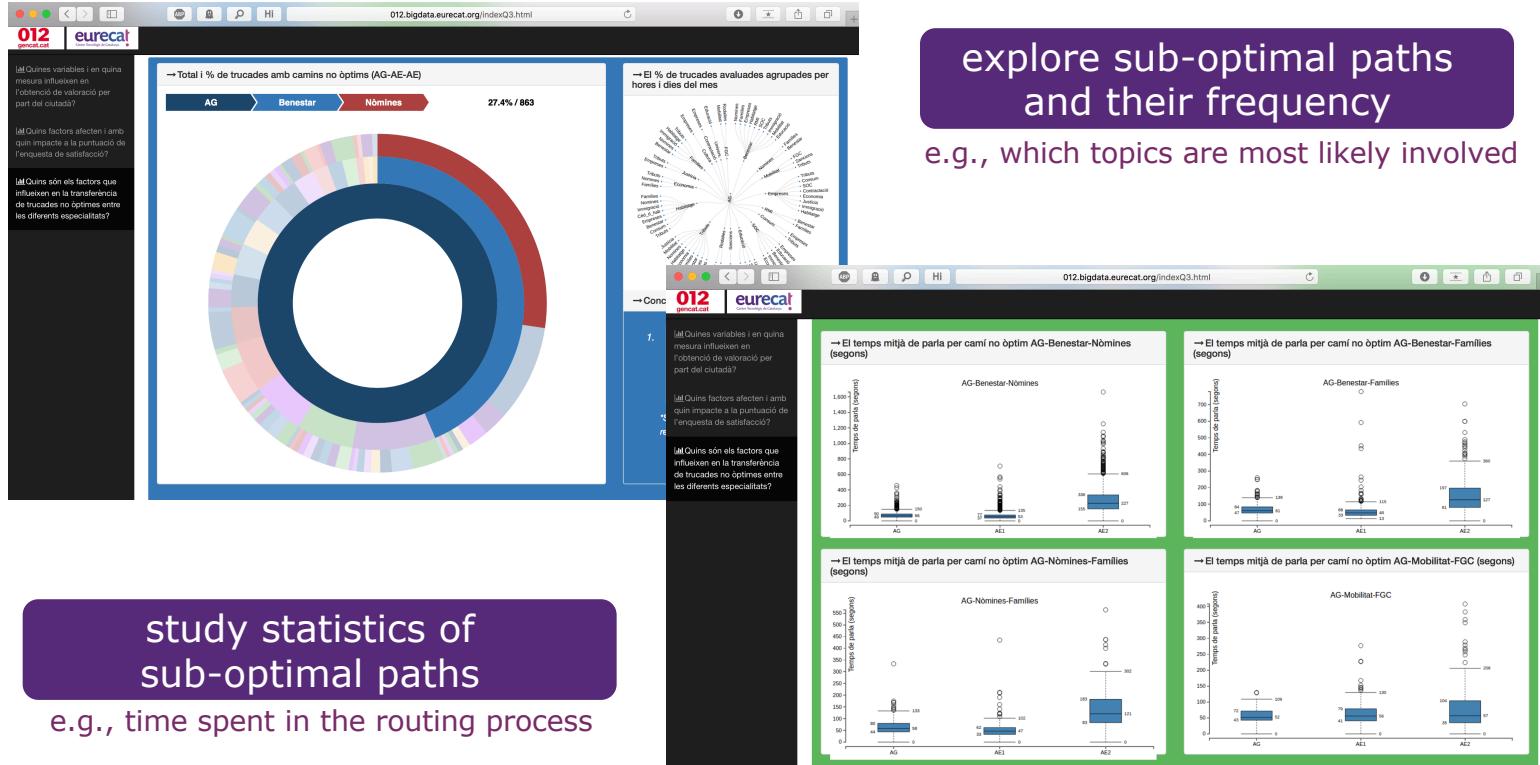
what are the factors that influence the rating the most?

what is the average talk time?



when bad ratings are more likely to happen?

012: Kibana dashboards



012: conclusions



Exploratory analysis

Statistical characterization of dataset



Dashboard

Inspect call stats, common routes, etc.



Outcome and informed actions

Internal optimizations cut the costs (18k€/year)



Project duration: 3 months

from handover of dataset to final reporting

LAVANGUARDIA | Vida



Al Minuto Internacional Política Opinión Vida Deportes Economía Local Gente Cultura Sucesos

Temas

Vida

Natural

Big Bang

Tecnología

Salud

Qué estudiar

Gente con alma

Junior Report

Bienestar

Más+

CAT-CATALUÑA 012

Generalitat aplica 'big data' para mejorar calidad y eficiencia servicio 012

f Comparte en Facebook

Twitter Comparte en Twitter



04/05/2017 12:12

Barcelona, 4 may (EFE).- La Generalitat de Cataluña ha puesto en marcha hoy un proyecto tecnológico de 'Big Data' que permitirá determinar los principales factores que inciden en la calidad del servicio de atención telefónica 012.

"Con este proyecto se demuestra el valor que genera la apertura y compartición de datos en las organizaciones, ya que generan un nuevo conocimiento que podemos usar para tomar decisiones", ha sentenciado el director general de Atención Ciudadana, Jordi Graells.

La iniciativa, en la que ha colaborado Eurecat, ha consistido en analizar a través de técnicas de analítica avanzada y de 'Big Data'-grandes bases de datos- los registros de 694.763 llamadas al 012 entre el mes de enero y octubre del año pasado para valorar la puntuación de los usuarios.

Con este análisis, el gobierno catalán ha concluido que el tiempo de espera entre quien atiende la llamada y el agente especialista del ámbito es uno de los factores valorados más negativamente por los usuarios.

Asimismo, este proyecto nace con el objetivo de crear propuestas alternativas para mejorar esta transferencia de llamadas, que según la Generalitat, generaría un ahorro anual de 18.000 euros. EFE

Más noticias



Nos jugamos Cataluña

El pleno de la declaración de independencia, en streaming

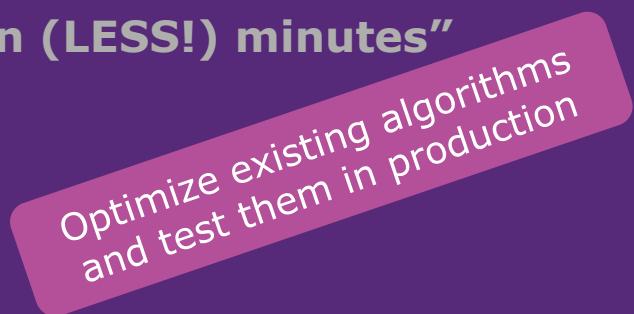
La hora de la verdad



Renuera tus básicos 'beauty' para este otoño por menos de 20 euros

Third story: **Glovo**

"Anything you want, delivered in (LESS!) minutes"



Optimize existing algorithms
and test them in production

Glovo

"Anything you want, delivered in minutes"



One of the coolest startup in town

- Last-mile delivery (meals, groceries, etc.)
- Shipment time <30minutes
- Partnership with other businesses
- Fast growing platform (>250K users)

- Been in the game for a couple of years
- Rather large dataset (live + historical)
- Existing back-end
- Time to take advantage of data...

Glovo: optimizing notification algorithm



notification
time



project goal(s)
new order notification algorithm based on estimated preparation time



idea

The notification should occur at the right moment

caveats

notify
too early

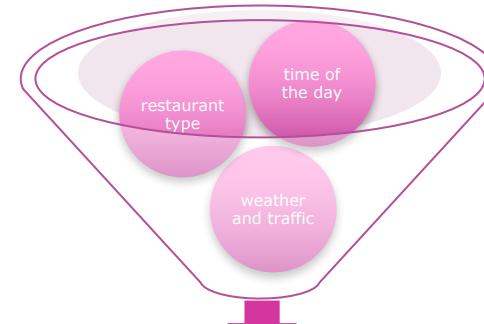
Courier arrives late
(cold) pizza waiting

notify
too late

Courier arrives early
higher waiting time

The right moment depends on the meal preparation time

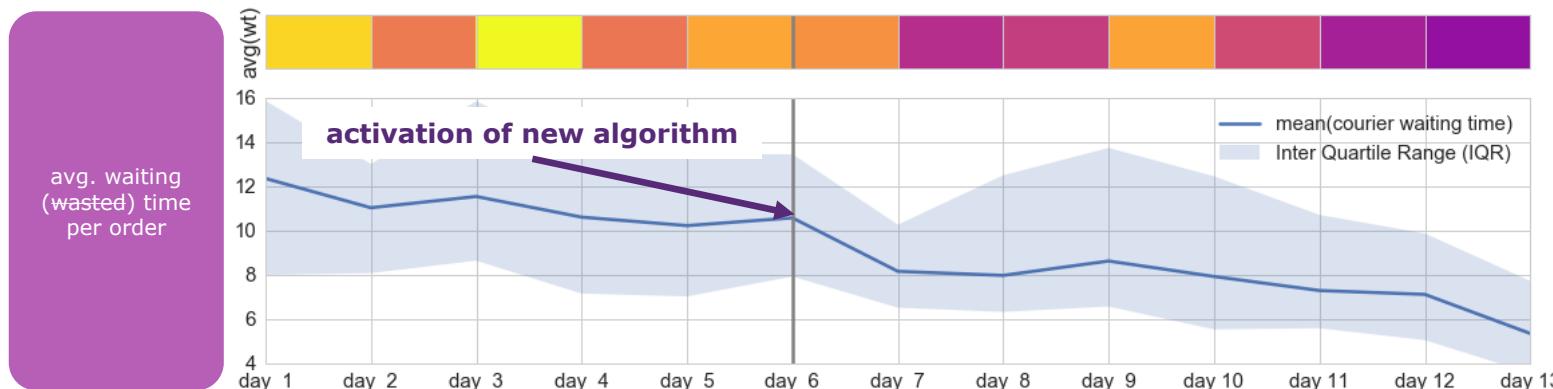
Glovo: optimizing notification algorithm



estimation of preparation time

Preparation time triggered notification

- Statistics of historical data + external factors
- Restaurant-based estimation of preparation time
- Calculate travel time to restaurant
- Notify restaurant at just the right time
- Minimize glover's waiting time
- Reduce overall delivery time



Glovo: conclusions



Exploratory analysis

Characterization of historical data



New notification algorithm

Statistical based



Courier waiting time decreased by ~40%

Major overall improvements



Project duration: 3 months

From exploratory analysis to deployment of new algorithm in production and tests

Small-term, agile projects

How-to make them work

- Clear/delimited goal (and be ready to modify it on the go)
- Communication is the key (product owner always in the loop)
- Fast prototyping process required
- The right tools: many out-the-shelf solutions would do
- The right team: senior profiles & cross functionality for the win

“Size matters not”



- Master Yoda

GRÀCIES THANK YOU

Pierdomenico Fiadino, PhD
pierdomenico.fiadino@eurecat.org
<http://fiadino.org>

info@eurecat.org
www.eurecat.org



[www.twitter.com/eurecat_news](https://twitter.com/eurecat_news)



[www.facebook.com/Eurecatorg](https://facebook.com/Eurecatorg)