Who goes there?
Profiling audience street by street with Telco data

Kael Huerta
Data Scientist
kael.huerta@toptal.com
@kaelhuerta
Today’s plan

- Introduction
- Dataset
- Models
- Final data product
The problem

- Telcos do not know their customers
The dataset

- SQL database
- Hadoop cluster
- Flat files
The dataset

- **SQL database**
  - Hadoop cluster
  - Flat files
The dataset

- SQL database
- Hadoop cluster
- Flat files
The dataset

- SQL database
- Hadoop cluster
- Flat files
The models

- Interest Profiling
- Scoring Segmentation
- Gender Prediction
- Age Prediction
Interest Profiling

Hire the top 3% of freelance talent
Interest Profiling

Open Catalog of Web Domains

Hire the top 3% of freelance talent
Interest Profiling
Interest Profiling
Interest Profiling

Life tip:
Research for what is done already
Scoring Segmentation
Scoring Segmentation

Fuzzy C-Means from e1071
Scoring Segmentation

Life tip:
Often, a *good-enough* model is enough
Gender Prediction
Gender Prediction

Logistic Lasso Regression from glmnet
Gender Prediction
Gender Prediction

Tech tip: `data.table`
Gender Prediction

Tech tip:

```
data.table
set.index(dt, cols, ...)
```
Age Prediction
Age Prediction

Factor Analysis from FactoMineR
Age Prediction

Life tip:
Share relatable examples
Final Data Product

$ + $ + $ + 18^+ = $

Hire the top 3% of freelance talent

www.toptal.com
Final Data Product
Final Data Product
Final Data Product

Package osmar
Final Data Product

Package sp
Final Data Product

Package leaflet
Applications
Applications

Hire the top 3% of freelance talent

www.toptal.com
Applications

Hire the top 3% of freelance talent
Applications

Hire the top 3% of freelance talent
Applications
Reach out!

Kael Huerta
Data Scientist
kael.huerta@toptal.com
@kaelhuerta

Hire the top 3% of freelance talent