Crime Patterns and Urban Infrastructure around Schools

Jaqueline Silveira

Advisor: Prof. Dr. Afonso Paiva Neto

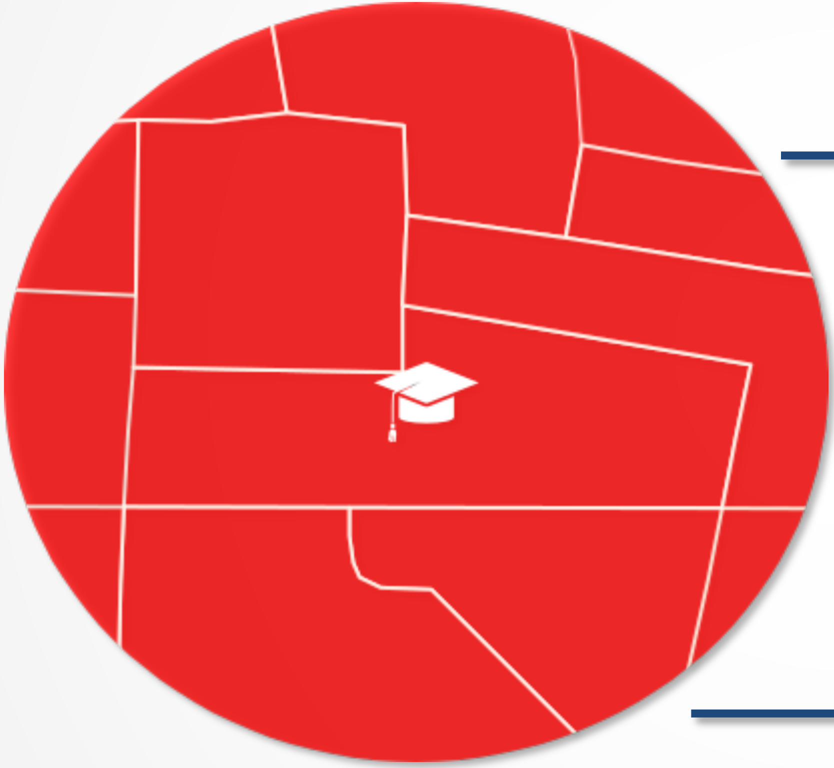
ICMC, University of São Paulo, Brazil

2020

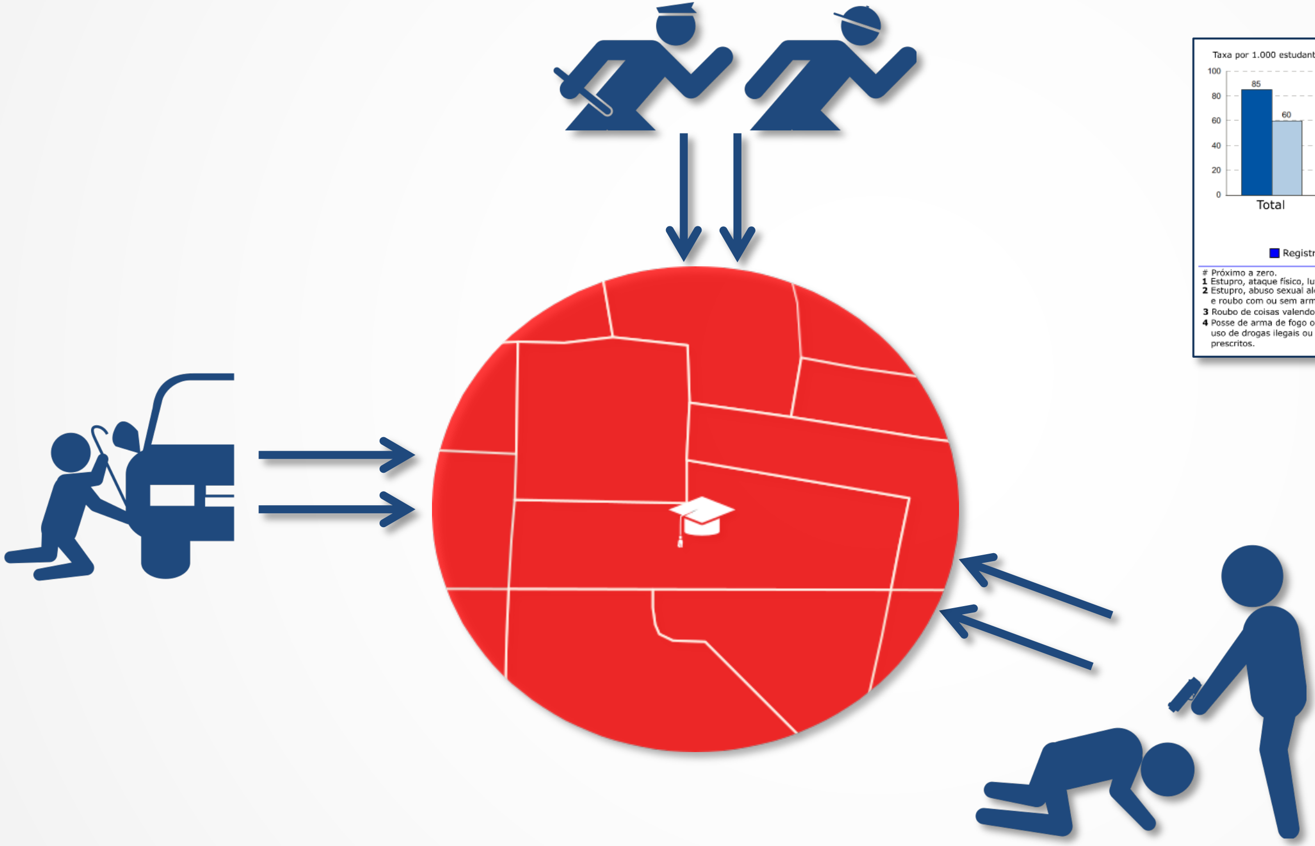
Outline:

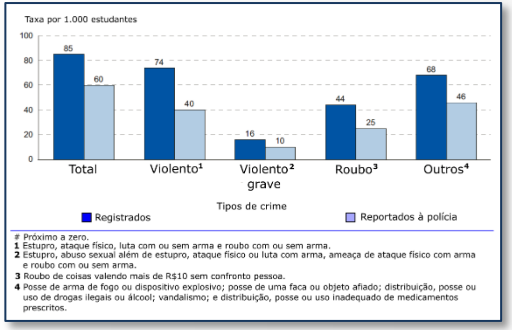
- Introduction
- Motivation
- TensorAnalyzer:
  - Pipeline
  - TensorAnalyzer versus feature vector
  - Case studies
- Mirante:
  - Case study
- Future directions

May crimes influence student's performance?



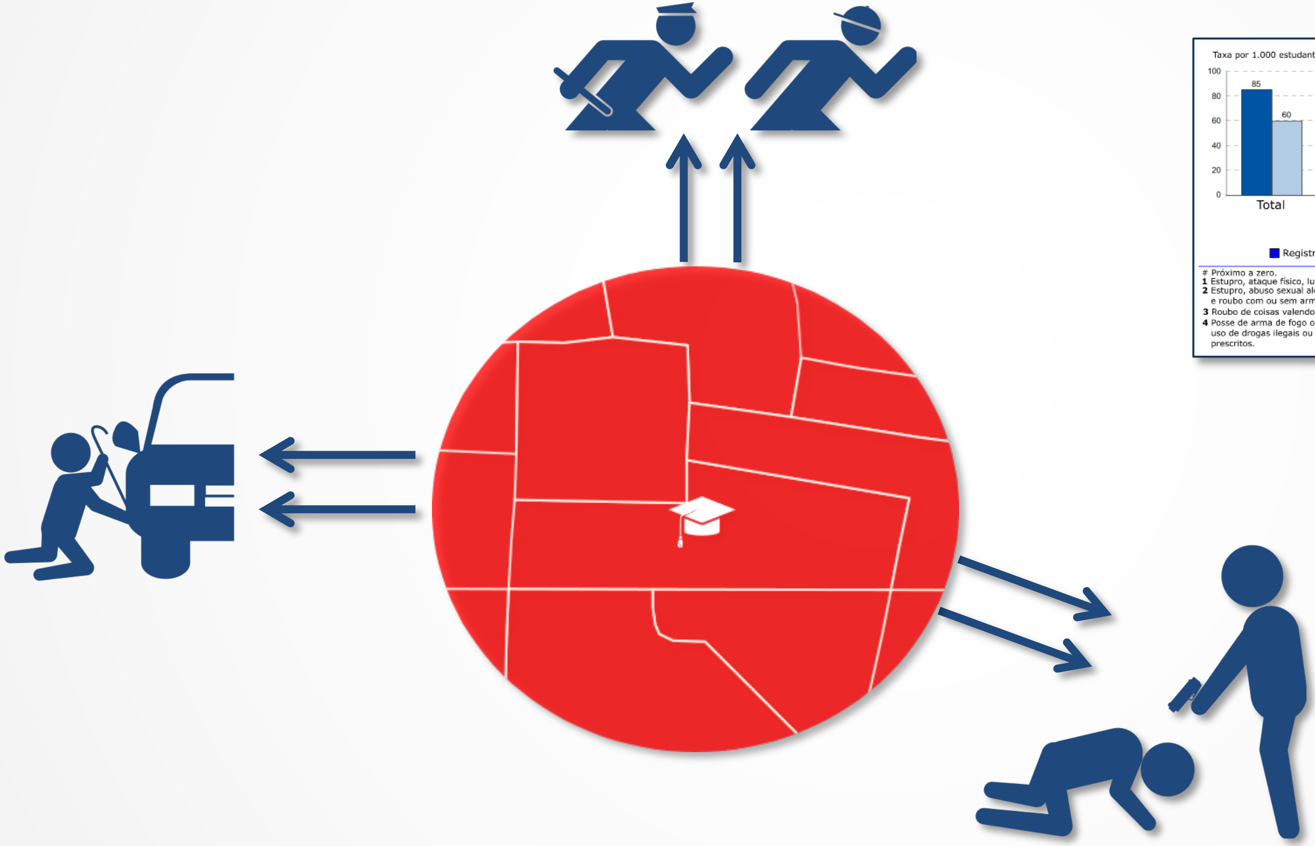
Motivation

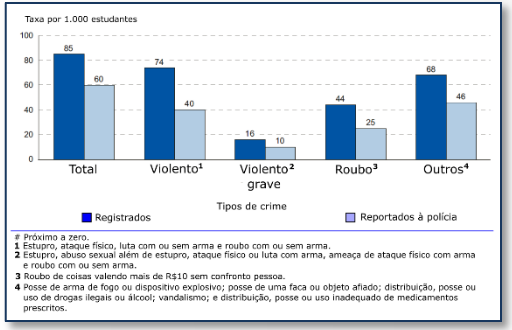




Roberts et al. (2012).

Motivation





Roberts et al. (2012).

Motivation

May the infrastructure influence crime increment?

Crime Studies

Infra structure

- *Bars*
- *Banks*
- *Schools*
- *Parks*
- *Etc*

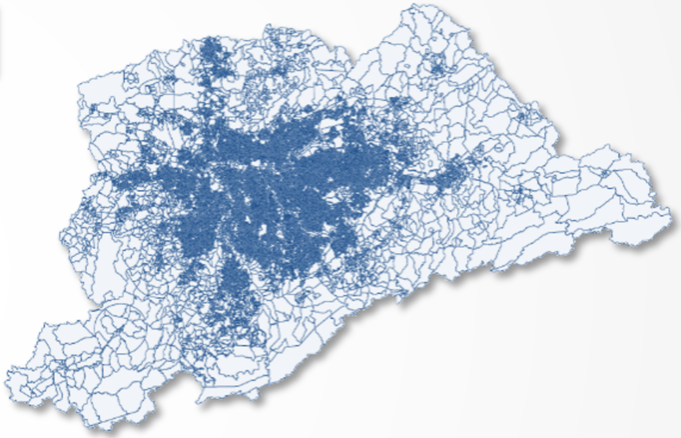
Socio Economic

- *Population*
- *Rent values*
- *Economic level*
- *Unemployment rate*
- *Etc.*

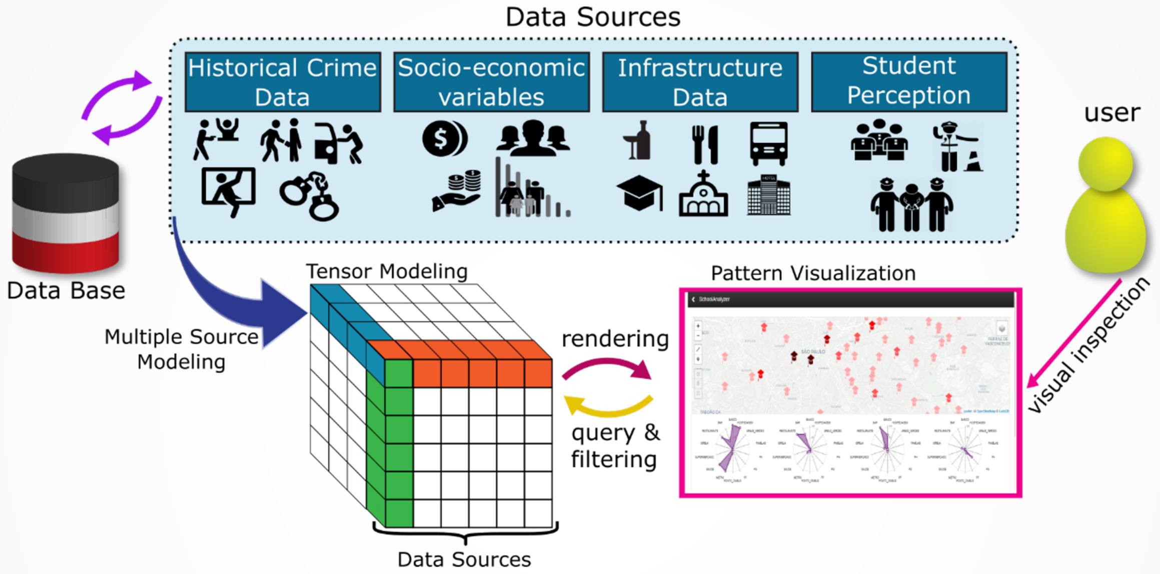
Historical Crimes

- *Passer-by*
- *Cargo theft*
- *Homicides*
- *Commercial Stablishment*
- *Etc.*

NEV
Núcleo de Estudos da Violência



Our Proposal:

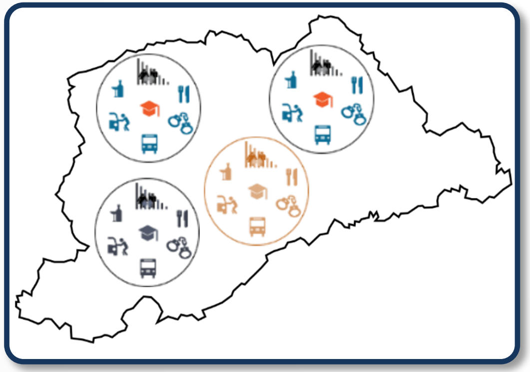


What are the relations between crime events and the other variables involved in the analysis?

What are the variables that most influence students' performance?

How to mathematically handle the multiple data sources in order to uncover patterns?

Our Proposal:



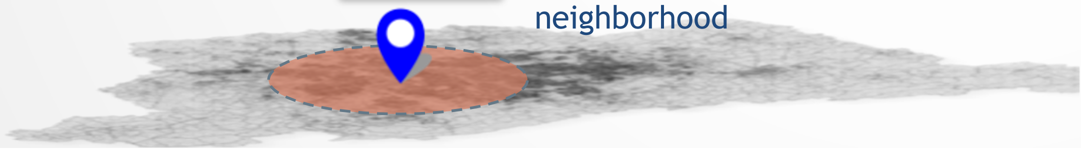
Infra Structure

Crimes

Urban Class

Crime Periods

School	Bus Stop	Passerby	AB	Morning	...	n
35290	15	13	1	0	...	n
35367	30	27	0	1	...	n
35155	12	8	1	1	...	n
35017	40	36	1	0	...	n
⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	m	m	m	m	...	m,n



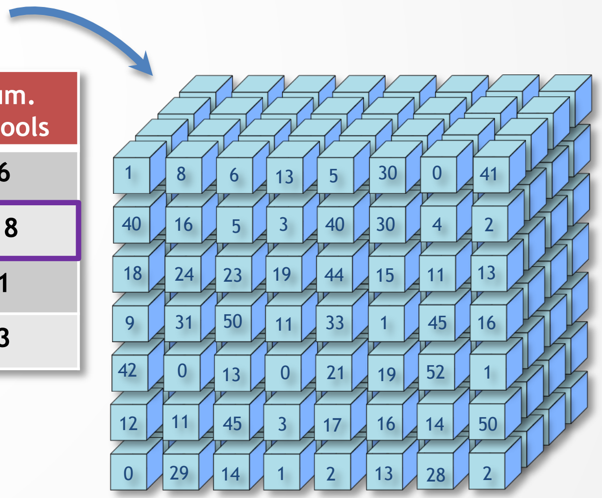
Our Proposal:

Bus Stop	Num. Schools
1-10	7
11-20	10
21-40	40
41-60	4
61-140	1

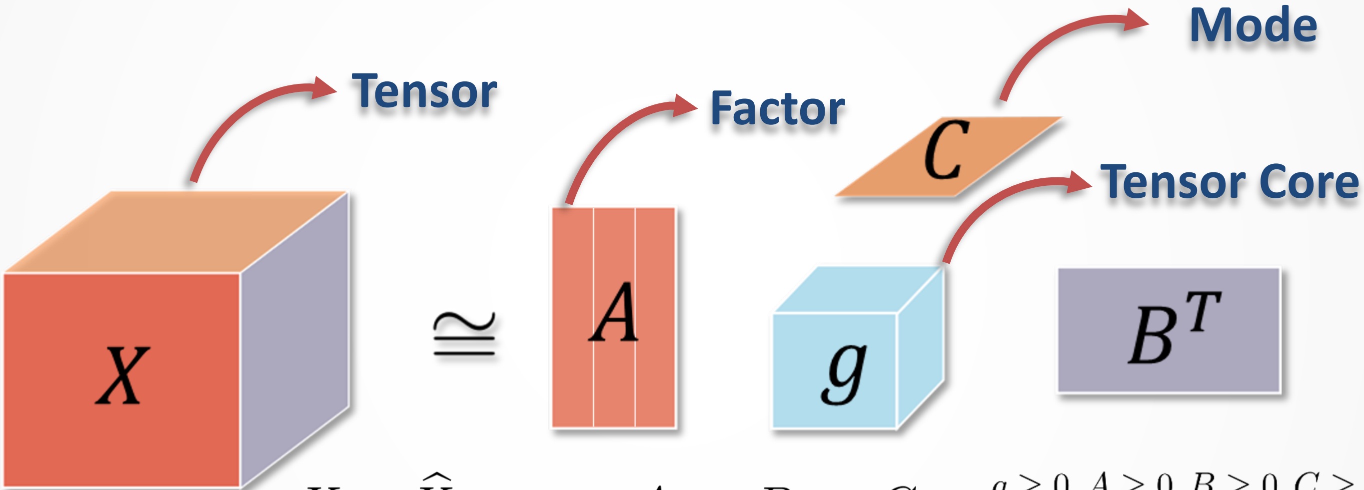
Est. Com.	Num. Schools
1-5	35
6-10	8
11-15	1
16-20	3
21-25	2

Urban Pattern	Num. Schools
AB	30
C	10
D	9
EFG	7
H	5

Crime Periods	Num. Schools
Dawn	6
Morning	18
Afternoon	1
Night	3



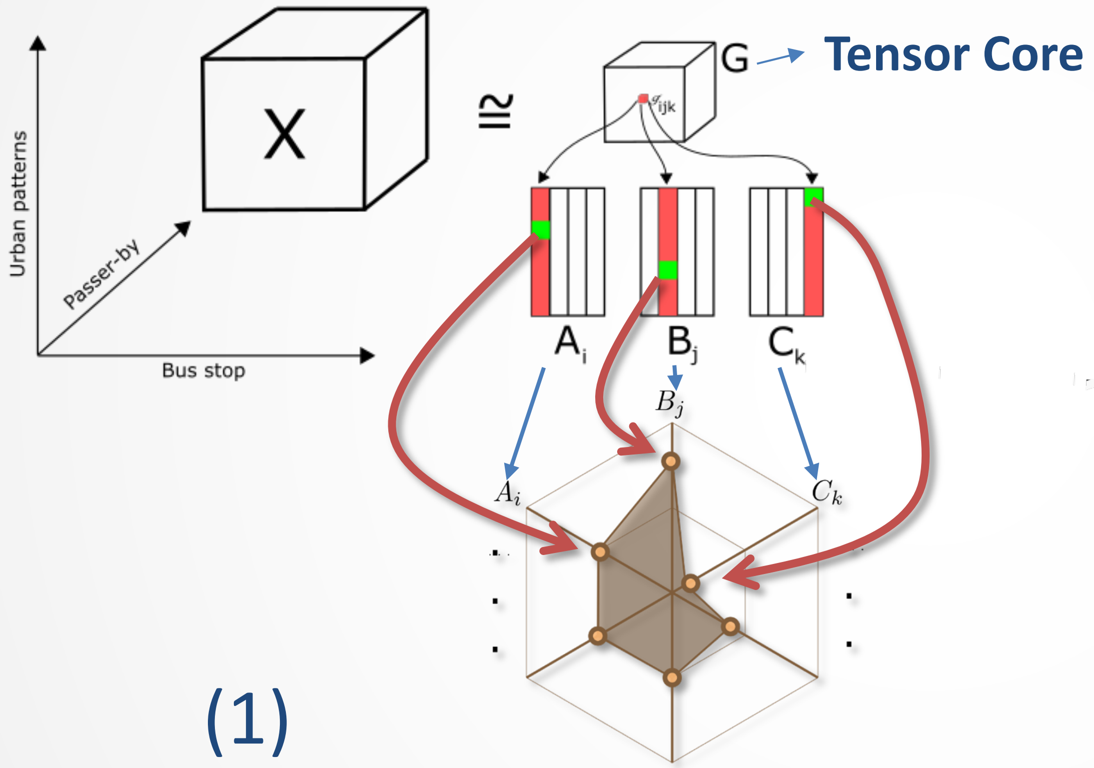
Tensor Tucker Decomposition



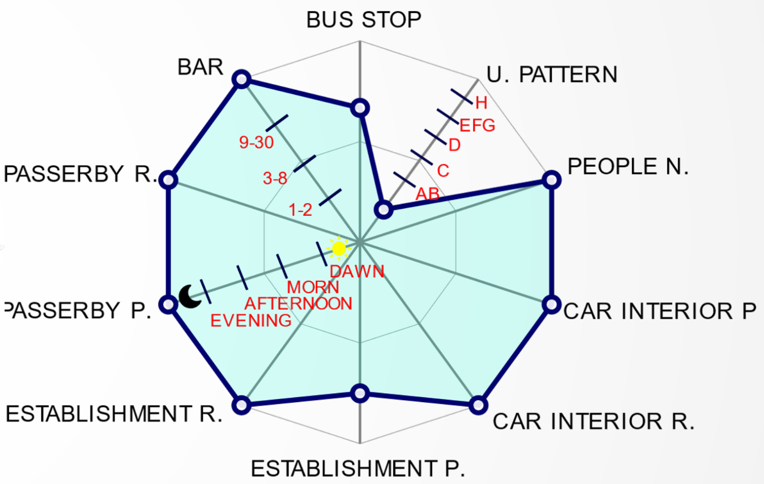
$$X \approx \widehat{X} = g \times_1 A \times_2 B \times_3 C$$

$$g \geq 0, A \geq 0, B \geq 0, C \geq 0$$

Our Proposal:

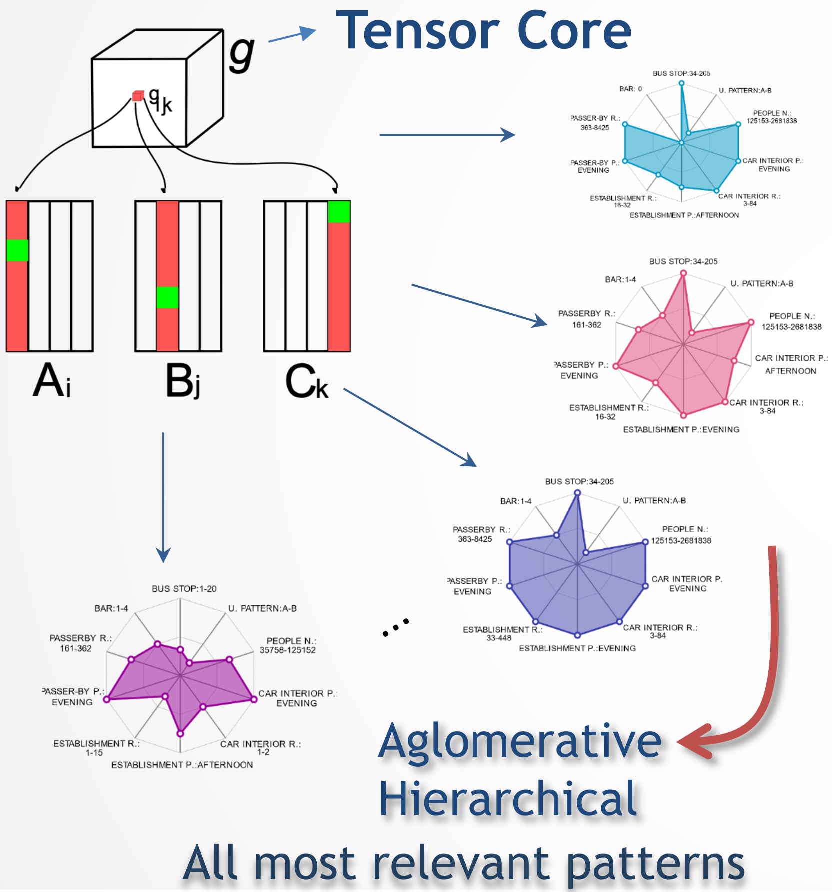


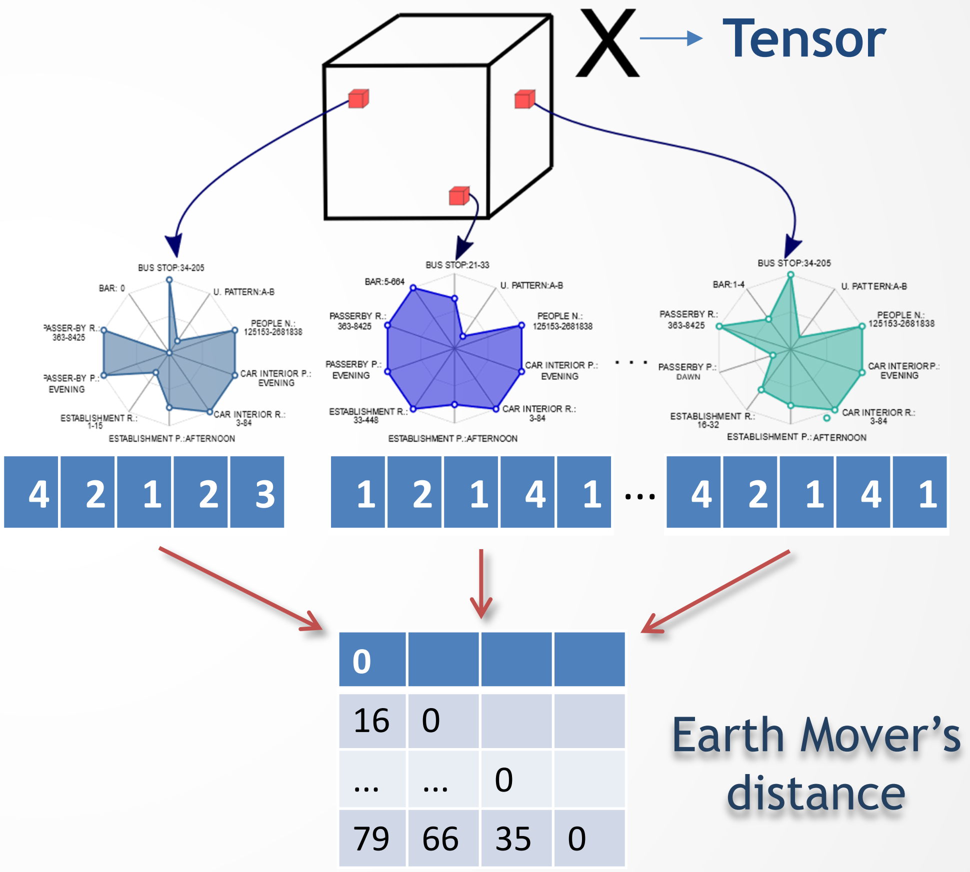
(1)



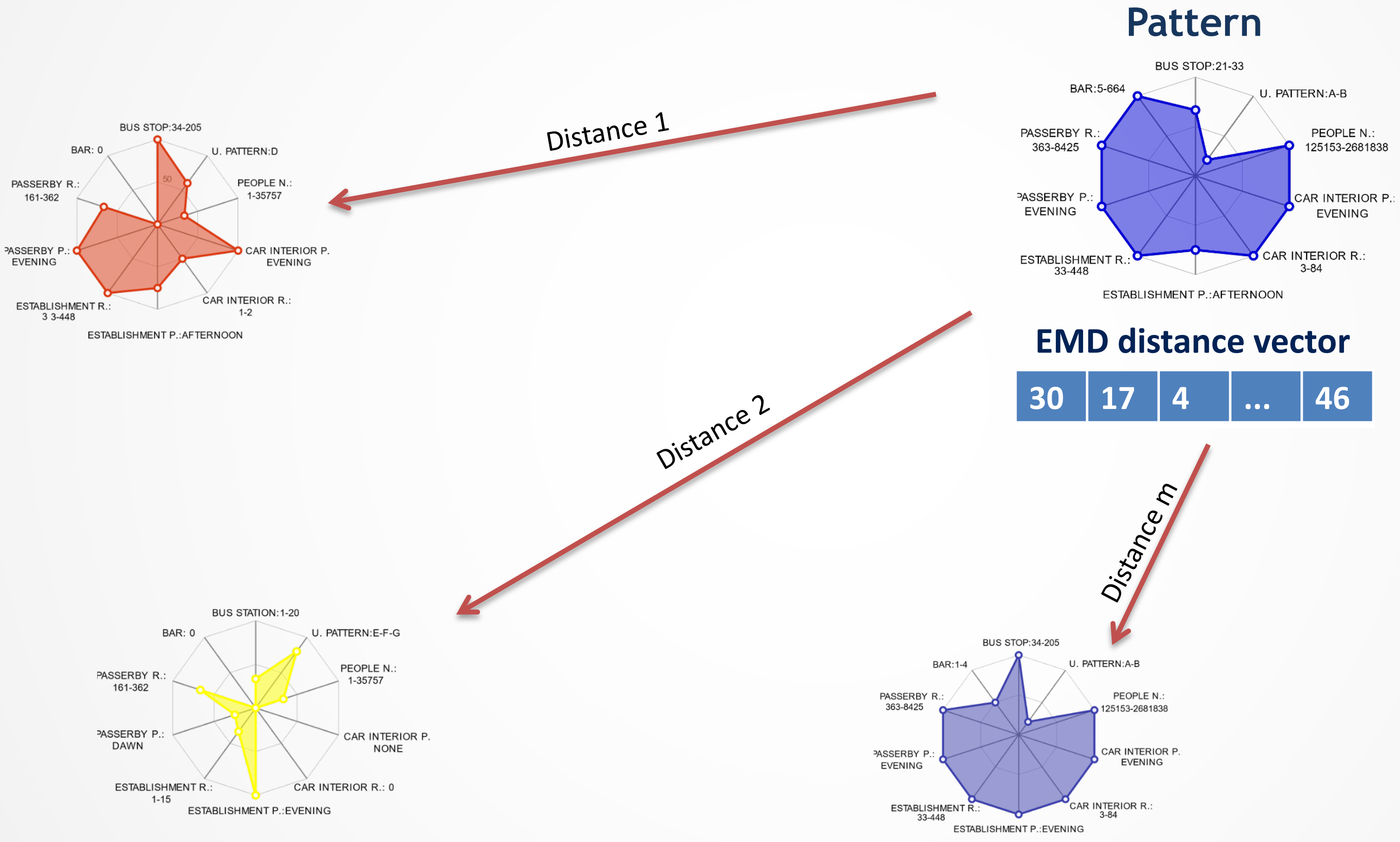
(2)

Pipeline Tensor - Approximation



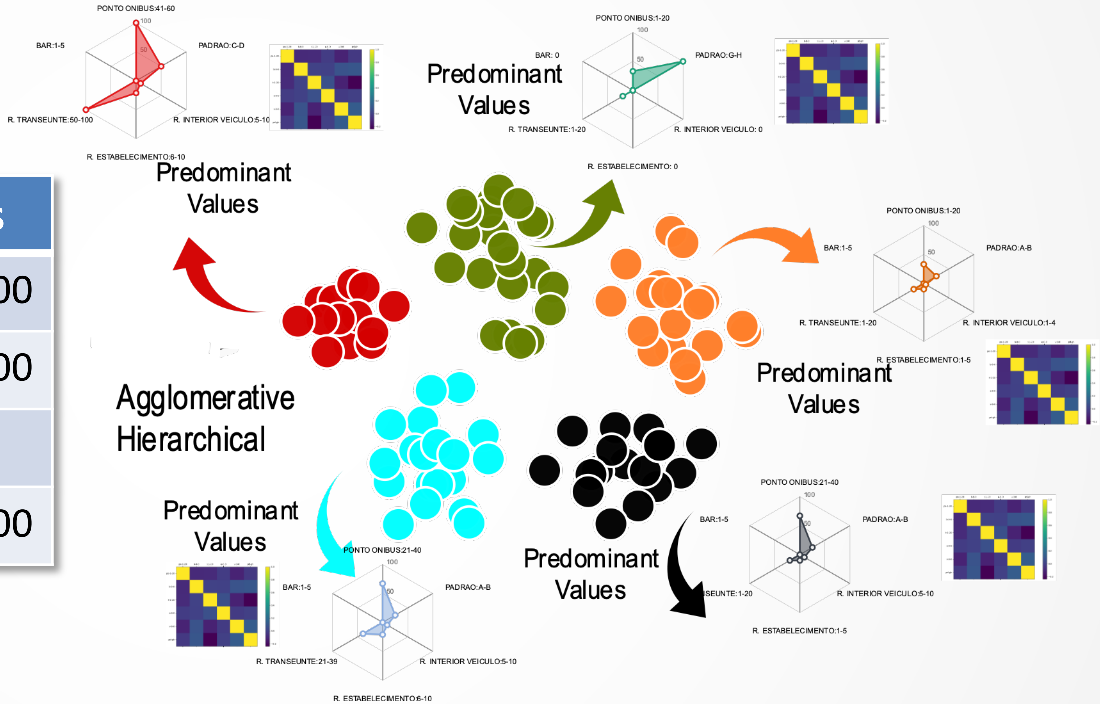


Pipeline Tensor - Approximation



Feature Vector

School	PO	Bar	…	Pass
E1	8	4	…	10000
E2	13	1	…	80000
…	…	…	…	…
En	22	0	…	329000

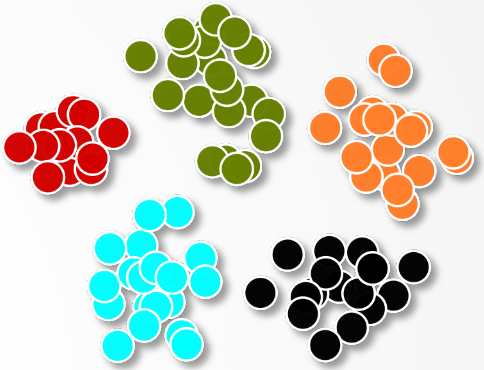


Results (2) - Synthetic data set (3.5k)

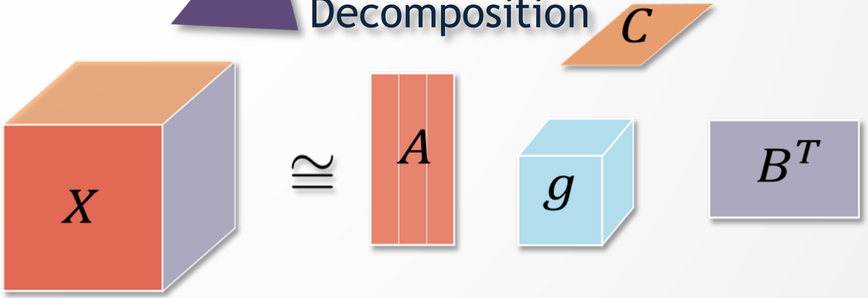
PO	Bar	Pass. R.	...	Vehi R.
11-20	3-9	30-60	...	5-10
...	...	...	...	...
40-50	20-30	90-100	...	15-21

School	PO	Bar	...	Pass
E1	8	4	...	10000
E2	13	1	...	80000
...	...	...	...	...
En	22	0	...	329000

Agglomerative Hierarchical Clustering



Tucker Decomposition



Results (2) - Feature Against Tensor

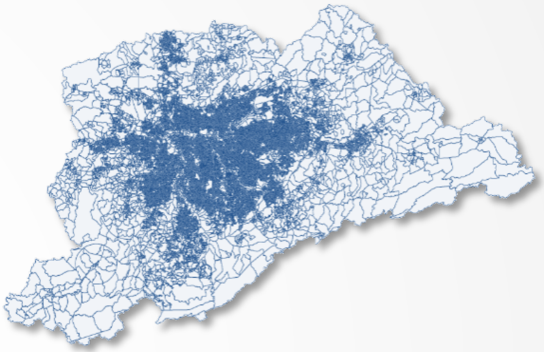
Clustering Performance Evaluation

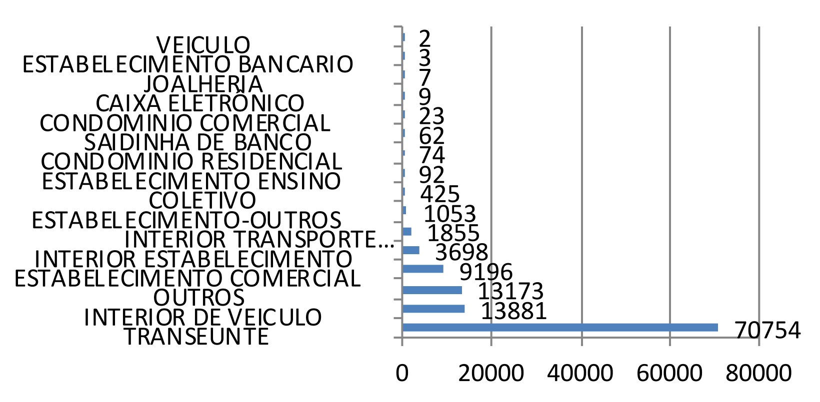
Metrics	Feature V.	Tucker
Fowlkes-Mallows scores	0.70091	0.94245
V-measure	0.80655	0.94491
Adjusted Rand index	0.61198	0.92791
Mutual Information	0.80531	0.94459

A Case Study in São Paulo

São Paulo Data Set

- Data set from 2011 to 2017
- Attributes:
  - **ANO_OCORR:** Year of occurrence.
  - **DATA_OCORRENCIA_BO:** Date of occurrence.
  - **HORA_OCORRENCIA_BO:** Hour of occurrence.
  - **NOME_DELEGACIA_CIRC:** Police station name
  - **RUBRICA:** Crime type (16 types)
  - **COD_SETOR:** Code of census block
  - **COORD_X:** lat
  - **COORD_Y:** lng



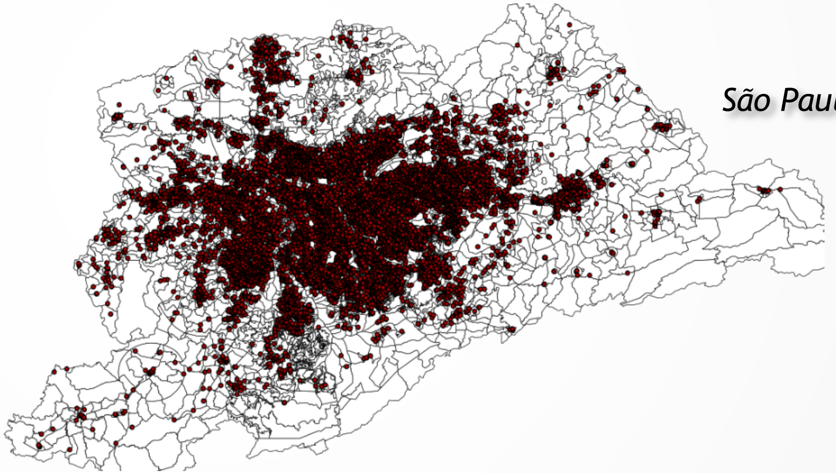


Is there any relationship between schools and criminality?

School Data Set

- Data set from 2016
- Attributes:
    - **CODESC:** school code.
    - **NOMEESC:** School Name.
    - **END_ESC:** School Address.
    - **CODSC2010:** Sector Code.
    - **COD_DEP:** Type of School.
    - **COORD_X:** lat
    - **COORD_Y:** lng

CEM
Centro de Estudos da Metrópole



São Paulo has ***11k*** *Schools*

Data Set:

NEV — Núcleo de Estudos da Violência

CEM — Centro de Estudos da Metrópole

PREFEITURA DE SÃO PAULO

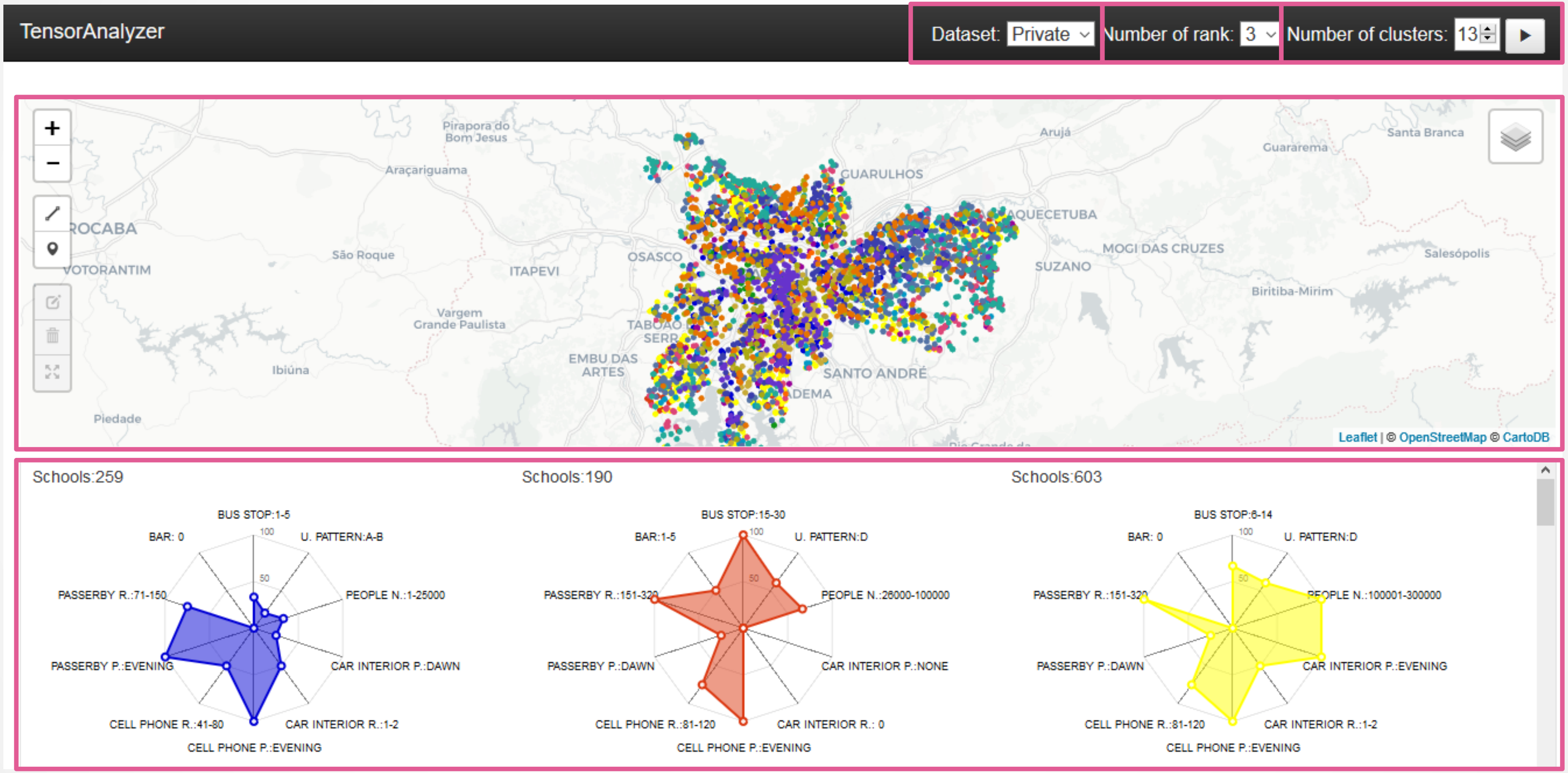
Crimes
- *Passerby*
- *Cell Phone*
- *Car Robbery*

Urban Pattern
- *A-B (High)*
- *C (Mid-High)*
- *D (Mid)*
- *E-F-G (Mid-Low)*
- *H (Low)*

Infra structure
- *Bus stops*
- *Bars*
- *People flow*

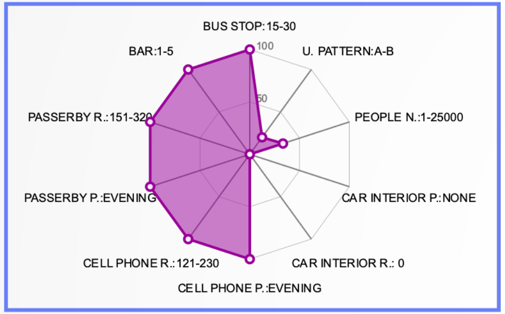
Crime Periods
- *Dawn*
- *Morning*
- *Afternoon*
- *Night*

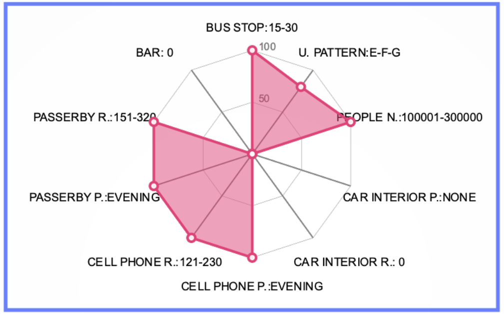
Results (1)

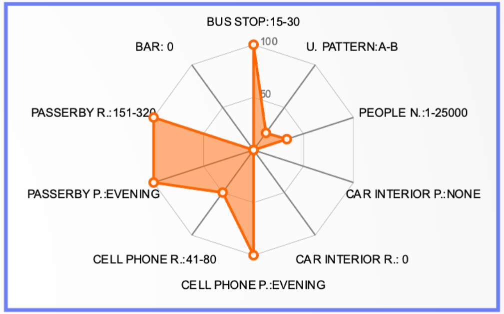

TensorAnalyzer
Dataset: Private
Number of rank: 3
Number of clusters: 13
Leaflet | © OpenStreetMap © CartoDB
Schools:259
BUS STOP:1-5
U. PATTERN:A-B
PEOPLE N.:1-25000
CAR INTERIOR P.:DAWN
CAR INTERIOR R.:1-2
CELL PHONE P.:EVENING
CELL PHONE R.:41-80
PASSERBY P.:EVENING
PASSERBY R.:71-150
BAR: 0
Schools:190
BUS STOP:15-30
U. PATTERN:D
PEOPLE N.:26000-100000
CAR INTERIOR P.:NONE
CAR INTERIOR R.: 0
CELL PHONE P.:EVENING
CELL PHONE R.:81-120
PASSERBY P.:DAWN
PASSERBY R.:151-320
BAR:1-5
Schools:603
BUS STOP:6-14
U. PATTERN:D
PEOPLE N.:100001-300000
CAR INTERIOR P.:EVENING
CAR INTERIOR R.:1-2
CELL PHONE P.:EVENING
CELL PHONE R.:81-120
PASSERBY P.:DAWN
PASSERBY R.:151-320
BAR: 0

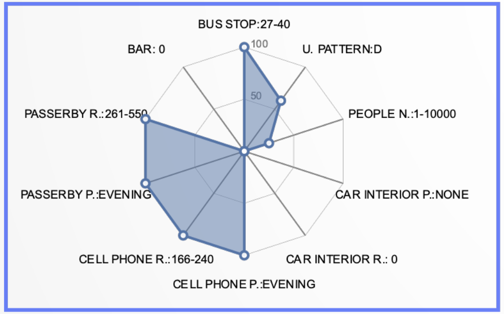
Results (2)

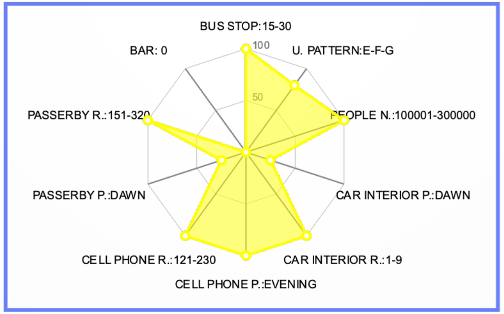
The relationship between crime events and the other variables.

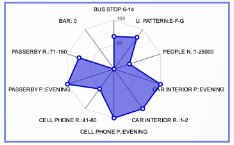


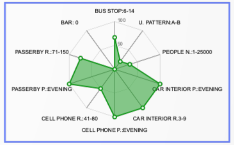


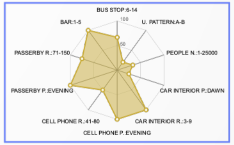












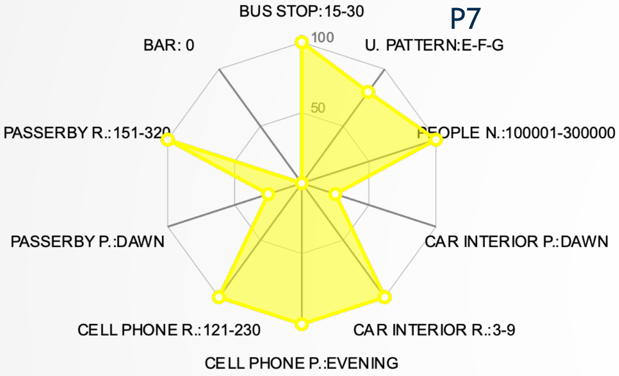
Results (2) - Regression

Ordinary Least Squares

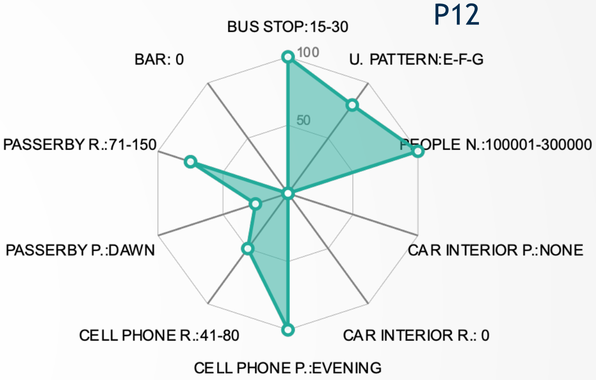
School	Infrastructure	Passerby	Cell Phone	Car Interior
Private	BUS STOP	5.2441	2.76	0.01
	BAR	3.0636	1.53	0.01
Town	BUS STOP	4.2692	2.4233	0.0108
	BAR	26.9055	13.8194	0.0925
State	BUS STOP	3.9218	2.3324	0.0148
	BAR	11.6920	5.2745	0.0758

Results (3) - Private Schools

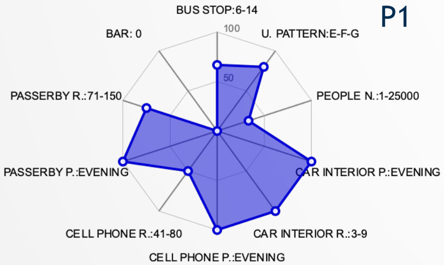
Variables that most influence students' performance.

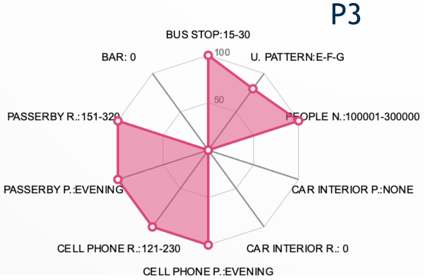


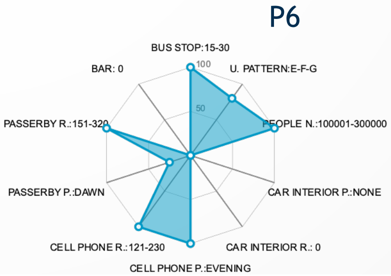
Pattern	EMAP	EMR	ENEM
P1	96.23	2.99	546
P3	93.65	3.61	554
P6	93.29	1.73	550
P7	93.30	3.78	537
P11	93.4	3.41	548
P12	91.89	4.34	598

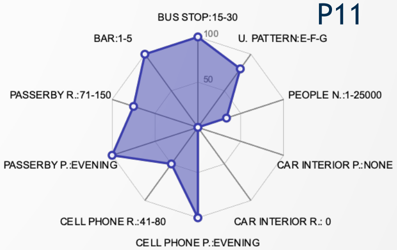


- **EMAP**: High school pass rate
- **EMR**: High school failed rate
- **ENEM**: National High School Exam





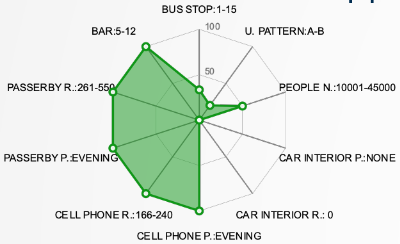




Results (3) - State Schools

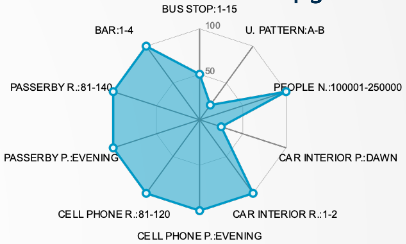
Variables that most influence students' performance.

P1



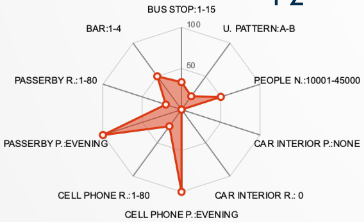
Pattern	EMAP	EMR	ENEM	IDEB_AI	IDEB_AF	TX_AI	TX_AF
P1	77.0	18.3	528	5.93	4.0	0.98	0.88
P2	78.1	17.0	577	5.91	4.4	0.98	0.90
P4	77.0	18.3	532	5.92	4.1	0.98	0.89
P5	73.0	19.6	492	5.52	3.9	0.97	0.88
P6	73.0	19.5	569	5.67	4.3	0.98	0.89
P7	74.0	18.2	560	5.50	4.2	0.97	0.89

P5

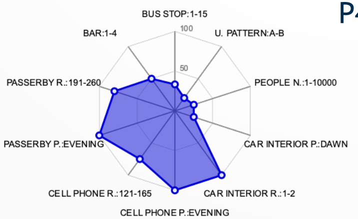


- **IDEB_AI:** Elementary schools educational development index in the initial years (1 to 5 years)
- **IDEB_AF**: Elementary schools educational development index in the final years (6 to 9 years)
- **TX_AI**: Internal student performance indicator in the initial years (1 to 5 years)
- **TX_AF**: Internal student performance indicator in the final years (6 to 9 years)

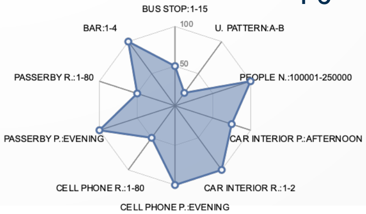
P2



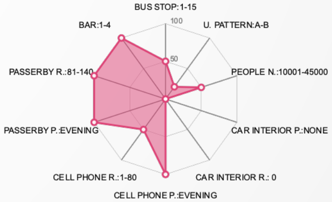
P4



P6



P7



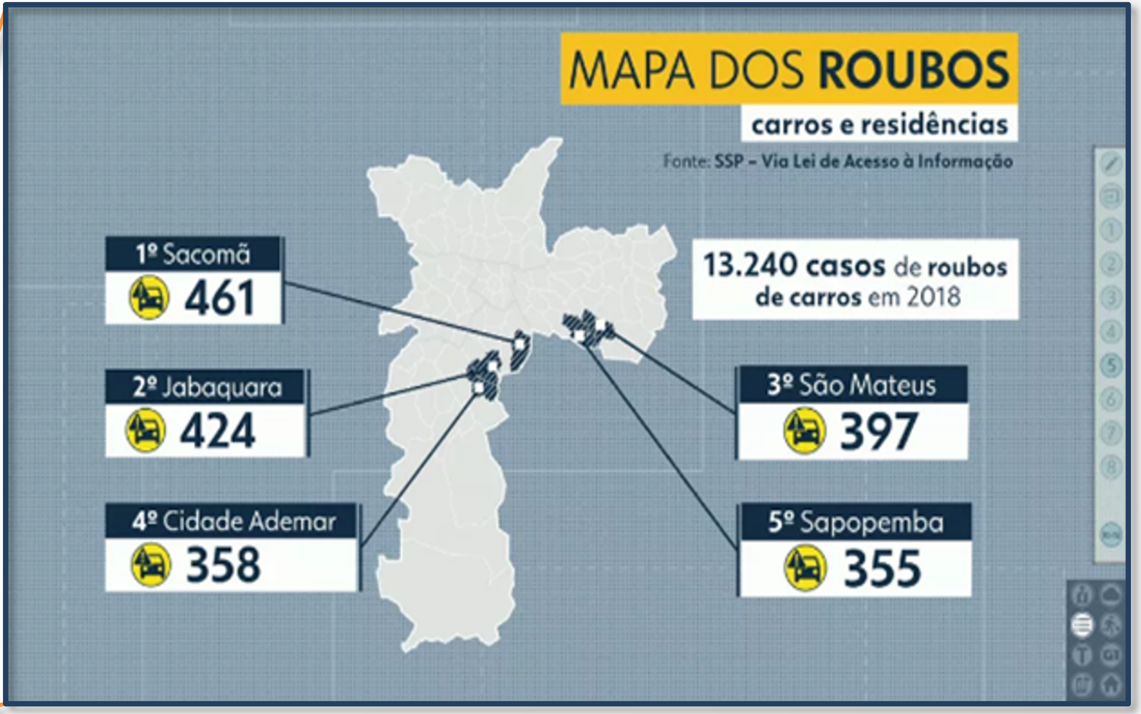
Sacomã é a região de SP com maior número de casos de roubos de carros e residências

Bairros da Zona Sul e da Zona Leste lideram o ranking de ocorrências na cidade.

Por Bom Dia SP — São Paulo

A cidade de **São Paulo** registrou em 2018 **13.240 mil casos de roubos de carros e 1.843 assaltos a residências**. O bairro de Sacomã, na Zona Sul , é líder nas duas ocorrências, de acordo com dados obtidos pelo Bom Dia SP com a Secretaria de Segurança Pública (SSP) via Lei de Acesso à Informação.

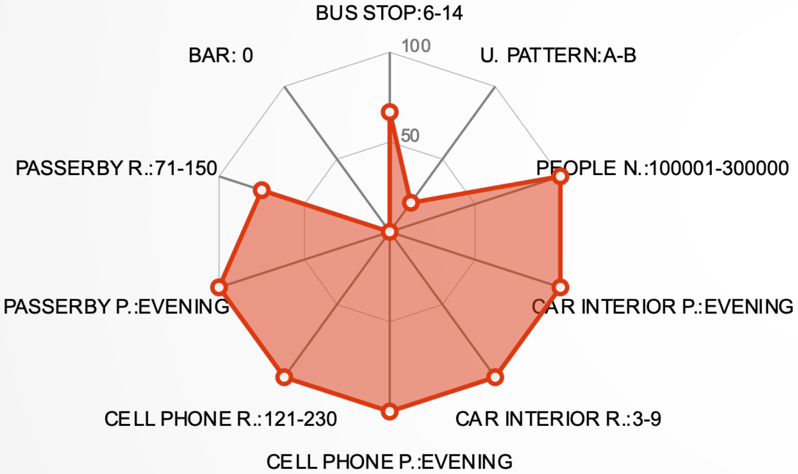
No ranking do roubo de veículos, **bairros da Zona Leste e da Zona Sul ocupam as primeiras posições**. Depois do Sacomã que tem 461 ocorrências, o Jabaquara está na 2ª posição com 424 casos. Já São Matheus aparece em 3º lugar com o registro de 397 roubos, em 4º está Cidade Ademar com 358 roubos e o em 5º lugar aparece Sapopemba com 355 casos.



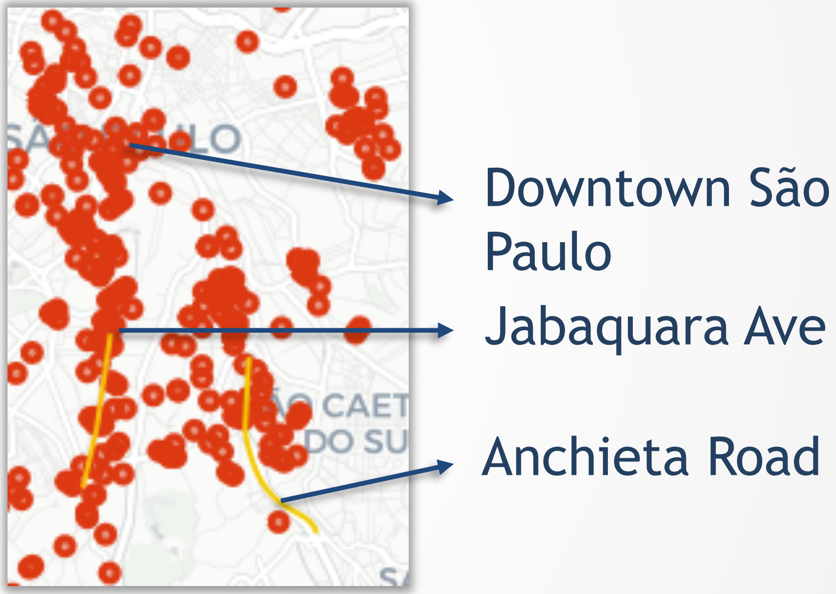
Fonte: g1.globo.com

Results (4)

“There are traffic jams over Anchieta road. Thus, the thieves take advantage that the cars are stopped, and they break the car glasses to steal bags, wallets, and others.”

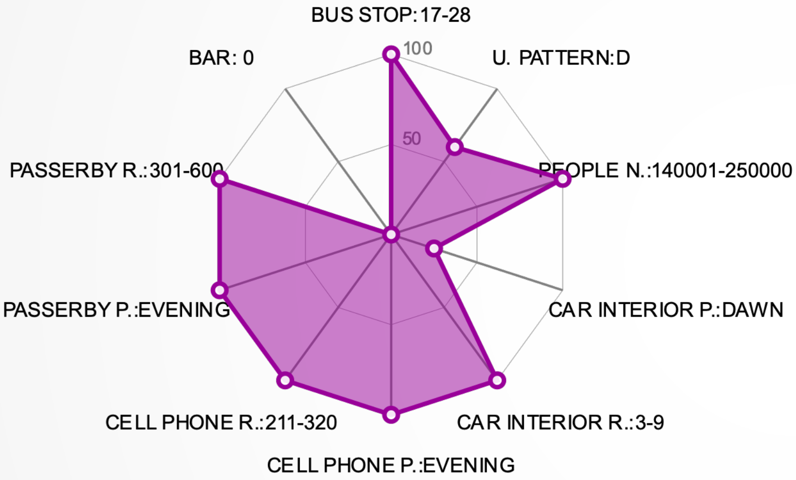


(1)



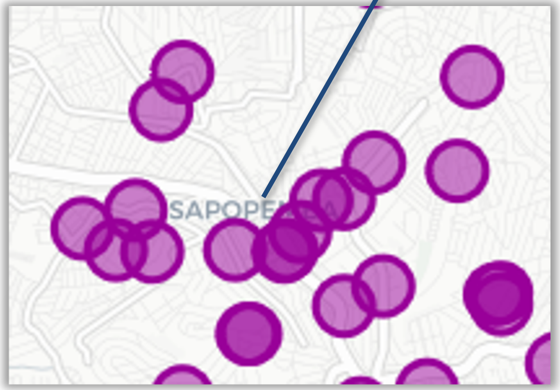
(2)

Results (4)

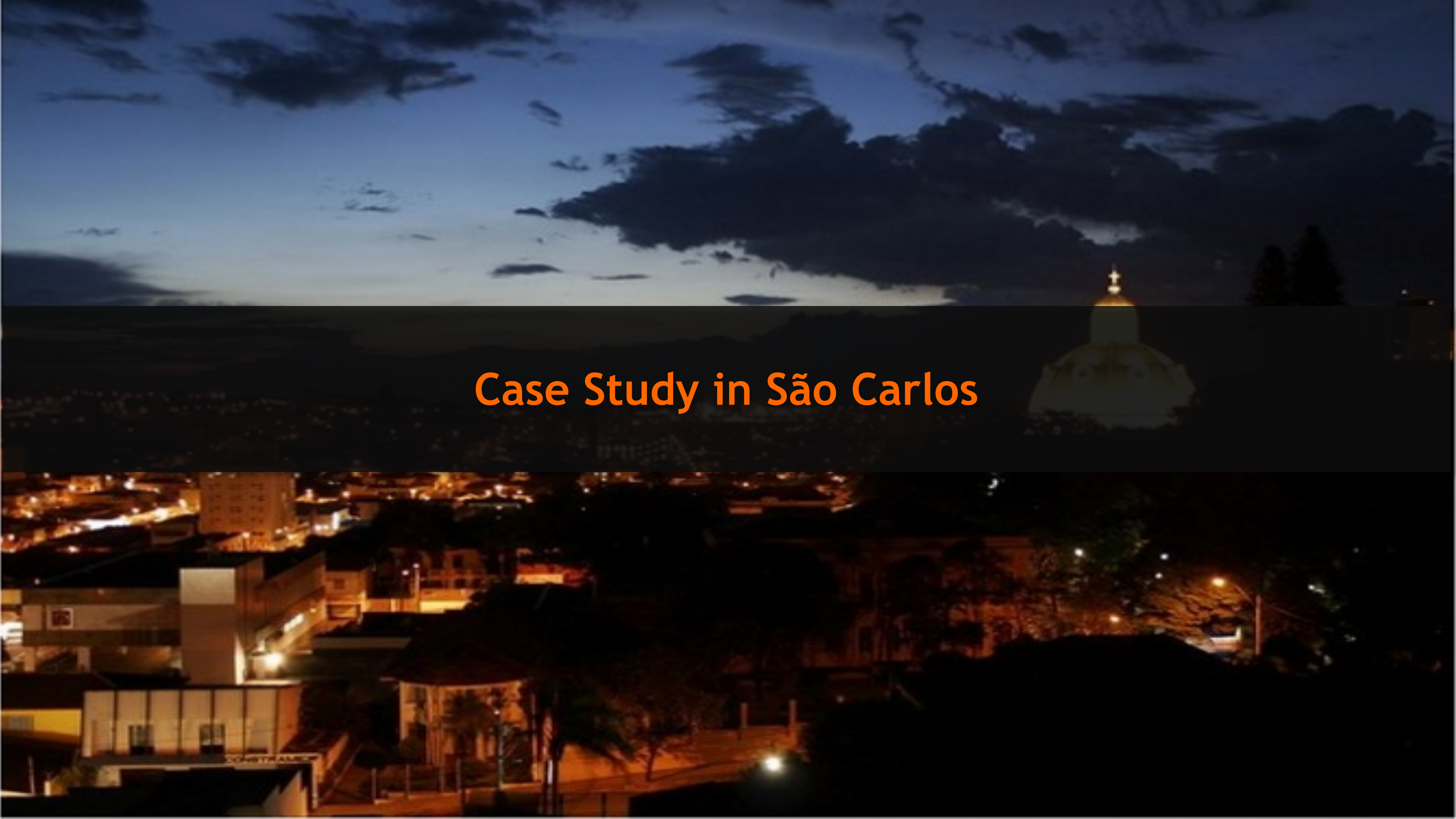


(1)

Sapopemba



(2)

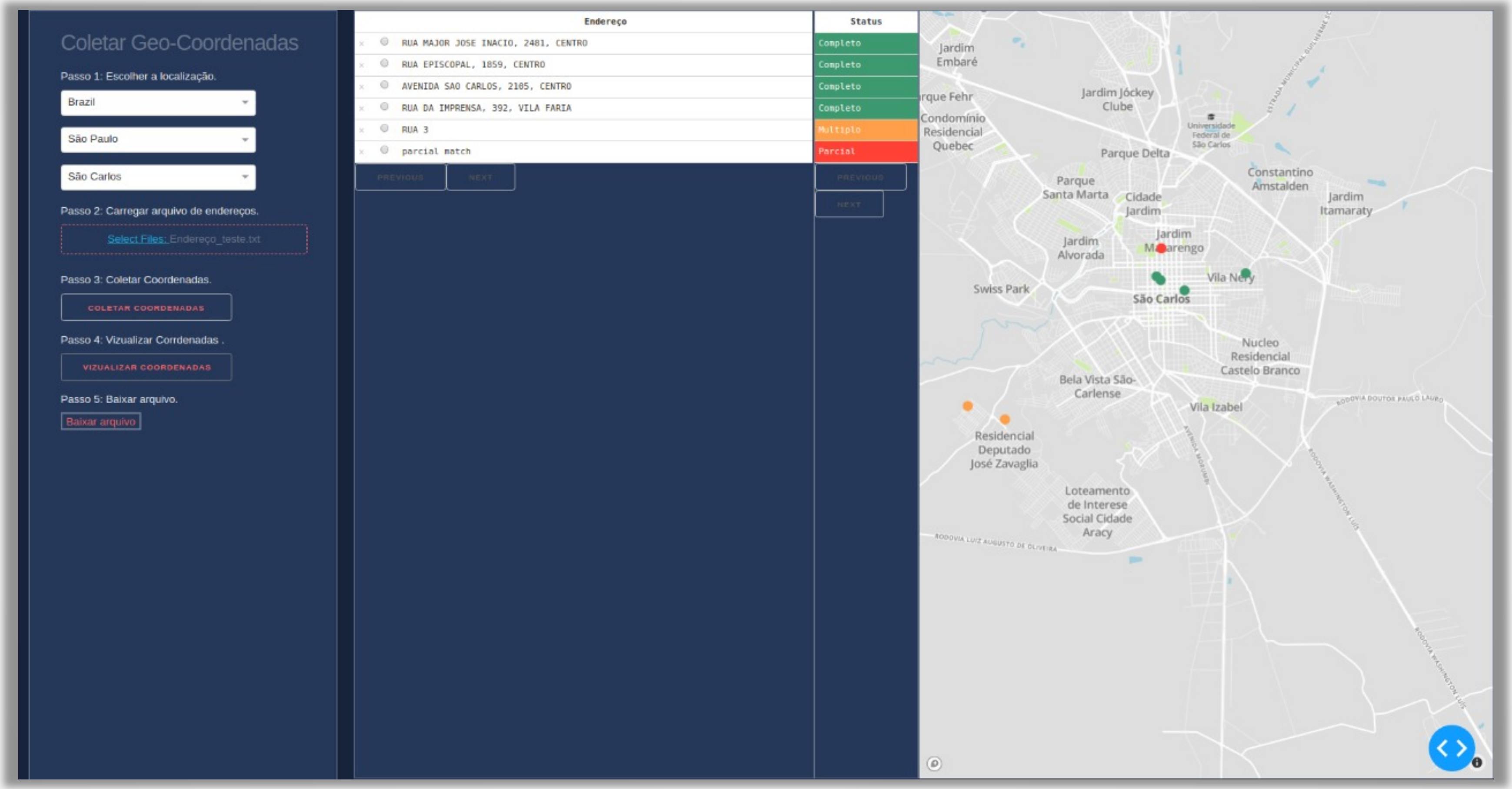


Case Study in São Carlos

Data Set

Data Base
Coordinate Collector
Google maps
Without Geo-Coordinates
With Geo-Coordinates
Before
After
Period
VEICULO
OUTROS
TRANSEUNTE
ESTABELECIMENTO COMERCIAL
CONDOMINIO COMERCIAL
ESTABELECIMENTO ENSINO
ESTABELECIMENTO-OUTROS
INTERIOR ESTABELECIMENTO
INTERIOR DE VEICULO
COLETIVO
INTERIOR TRANSPORTE COLETIVO
SAIDINHA DE BANCO
CAIXA ELETRÔNICO
Crimes Occurences
Schools
Municipal
Estadual
Particular
Outros

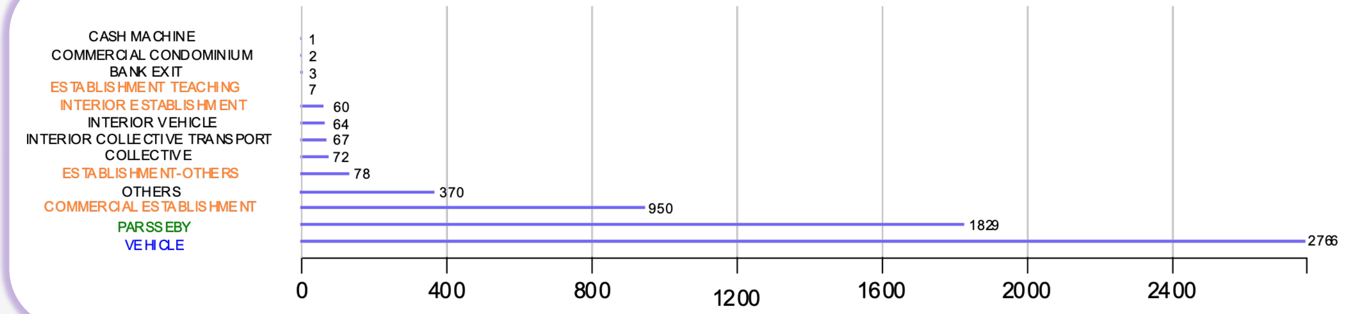
Data Set



Coletar Geo-Coordenadas
Passo 1: Escolher a localização.
Brazil
São Paulo
São Carlos
Passo 2: Carregar arquivo de endereços.
Select Files: Endereço_teste.txt
Passo 3: Coletar Coordenadas.
COLETAR COORDENADAS
Passo 4: Vizualizar Corrdenadas .
VIZUALIZAR COORDENADAS
Passo 5: Baixar arquivo.
Baixar arquivo
Endereço
RUA MAJOR JOSE INACIO, 2481, CENTRO
RUA EPISCOPAL, 1859, CENTRO
AVENIDA SAO CARLOS, 2105, CENTRO
RUA DA IMPRENSA, 392, VILA FARIA
RUA 3
parcial match
Status
Completo
Completo
Completo
Completo
Multiplo
Parcial
Jardim Embaré
Jardim Jóckey Clube
Condomínio Residencial Quebec
Parque Delta
Constantino Amstalden
Parque Santa Marta
Cidade Jardim
Jardim Itamaraty
Jardim Alvorada
Vila Nery
Swiss Park
São Carlos
Nucleo Residencial Castelo Branco
Bela Vista São-Carlense
Vila Izabel
Residencial Deputado José Zavaglia
Loteamento de Interese Social Cidade Aracy

Data Set

- Data set from 2014 to 2019
- Atributes:
  - **ANO:** Year of occurrence.
  - **DATA_OCORRENCIA_BO:** Date of occurrence.
  - **HORA_OCORRENCIA_BO:** Hour of occurrence.
  - **FLAGRANTE:** Flagrant
  - **CONDUTA:** Type of crime (13 types)
  - **LATITUDE:** lat
  - **LONGITUDE:** lng



Data Set:

Crimes

- *Passerby*
- *Commercial establishment*
- *Car Robbery*

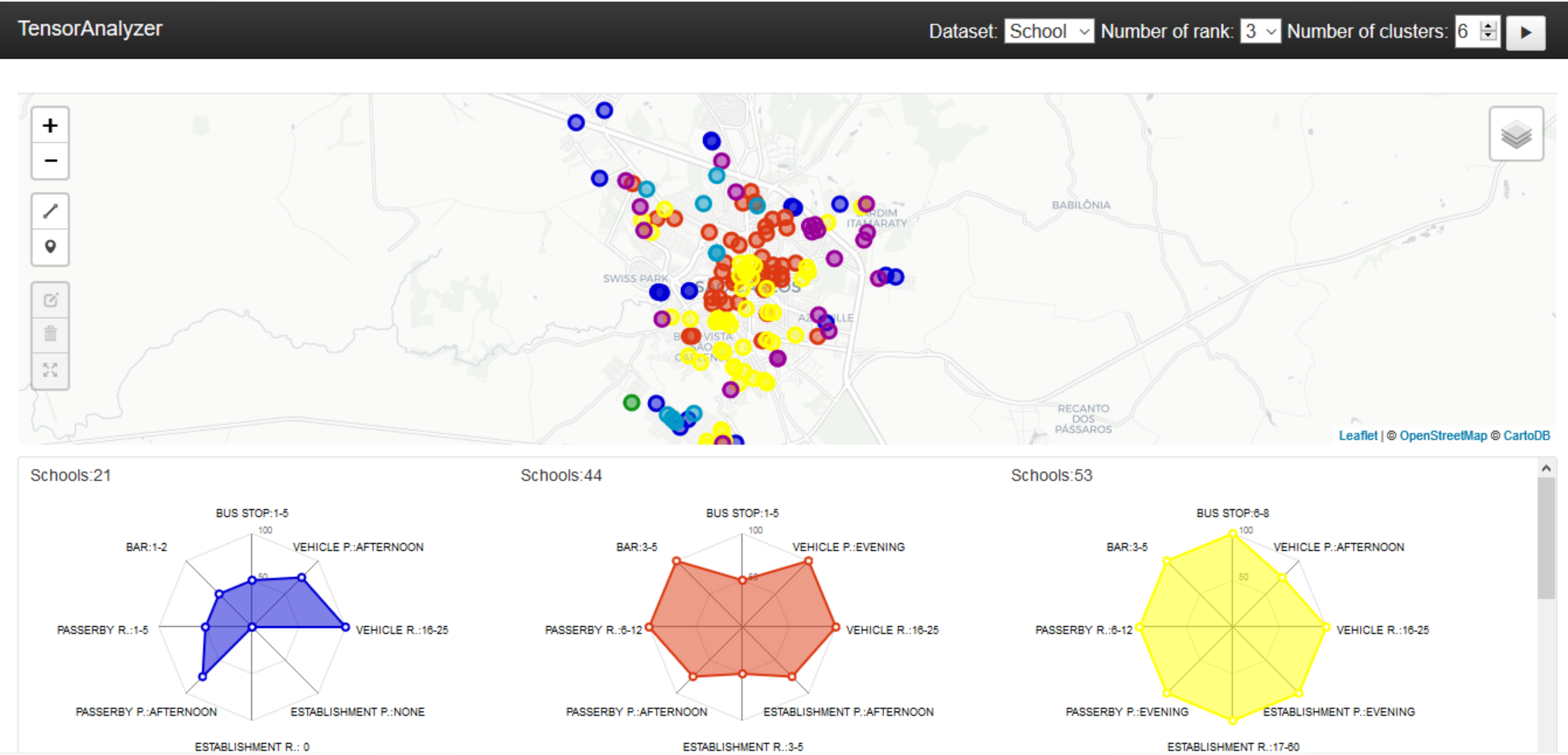
Crime Periods

- *Dawn*
- *Morning*
- *Afternoon*
- *Night*

Infra structure

- *Bus stops*
- *Bars*

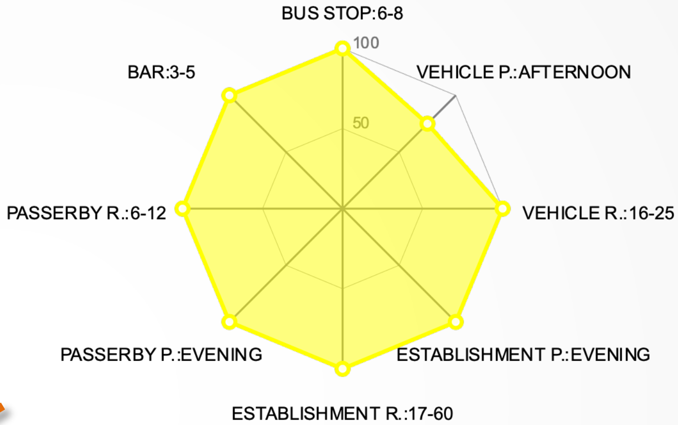
Results (1)



TensorAnalyzer
Dataset: School
Number of rank: 3
Number of clusters: 6
BABILÔNIA
SWISS PARK
RECANTO DOS PÁSSAROS
Leaflet | © OpenStreetMap © CartoDB
Schools:21
BUS STOP:1-5
VEHICLE P.:AFTERNOON
VEHICLE R.:16-25
ESTABLISHMENT P.:NONE
ESTABLISHMENT R.: 0
PASSERBY P.:AFTERNOON
PASSERBY R.:1-5
BAR:1-2
100
50
Schools:44
BUS STOP:1-5
VEHICLE P.:EVENING
VEHICLE R.:16-25
ESTABLISHMENT P.:AFTERNOON
ESTABLISHMENT R.:3-5
PASSERBY P.:AFTERNOON
PASSERBY R.:6-12
BAR:3-5
100
50
Schools:53
BUS STOP:6-8
VEHICLE P.:AFTERNOON
VEHICLE R.:16-25
ESTABLISHMENT P.:EVENING
ESTABLISHMENT R.:17-60
PASSERBY P.:EVENING
PASSERBY R.:6-12
BAR:3-5
100
50

Results (2)


R. São Seba
Google

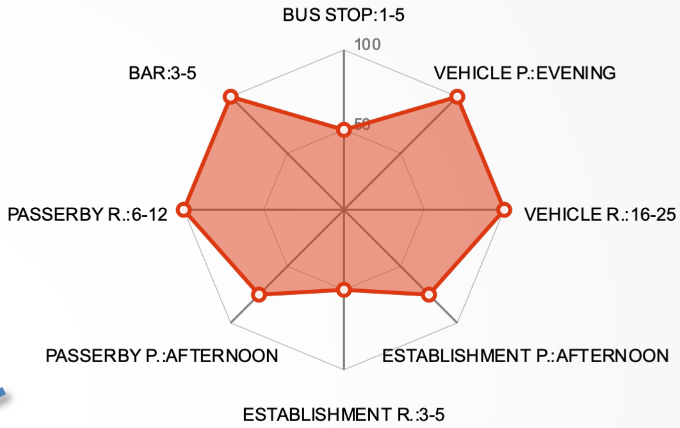

BUS STOP:6-8
100
50
VEHICLE P.:AFTERNOON
VEHICLE R.:16-25
ESTABLISHMENT P.:EVENING
ESTABLISHMENT R.:17-60
PASSERBY P.:EVENING
PASSERBY R.:6-12
BAR:3-5


Google


Google

Results (2)


R. Dona
Moreira
Google


BUS STOP:1-5
100
50
VEHICLE P.:EVENING
VEHICLE R.:16-25
ESTABLISHMENT P.:AFTERNOON
ESTABLISHMENT R.:3-5
PASSERBY P.:AFTERNOON
PASSERBY R.:6-12
BAR:3-5


PARE
Google

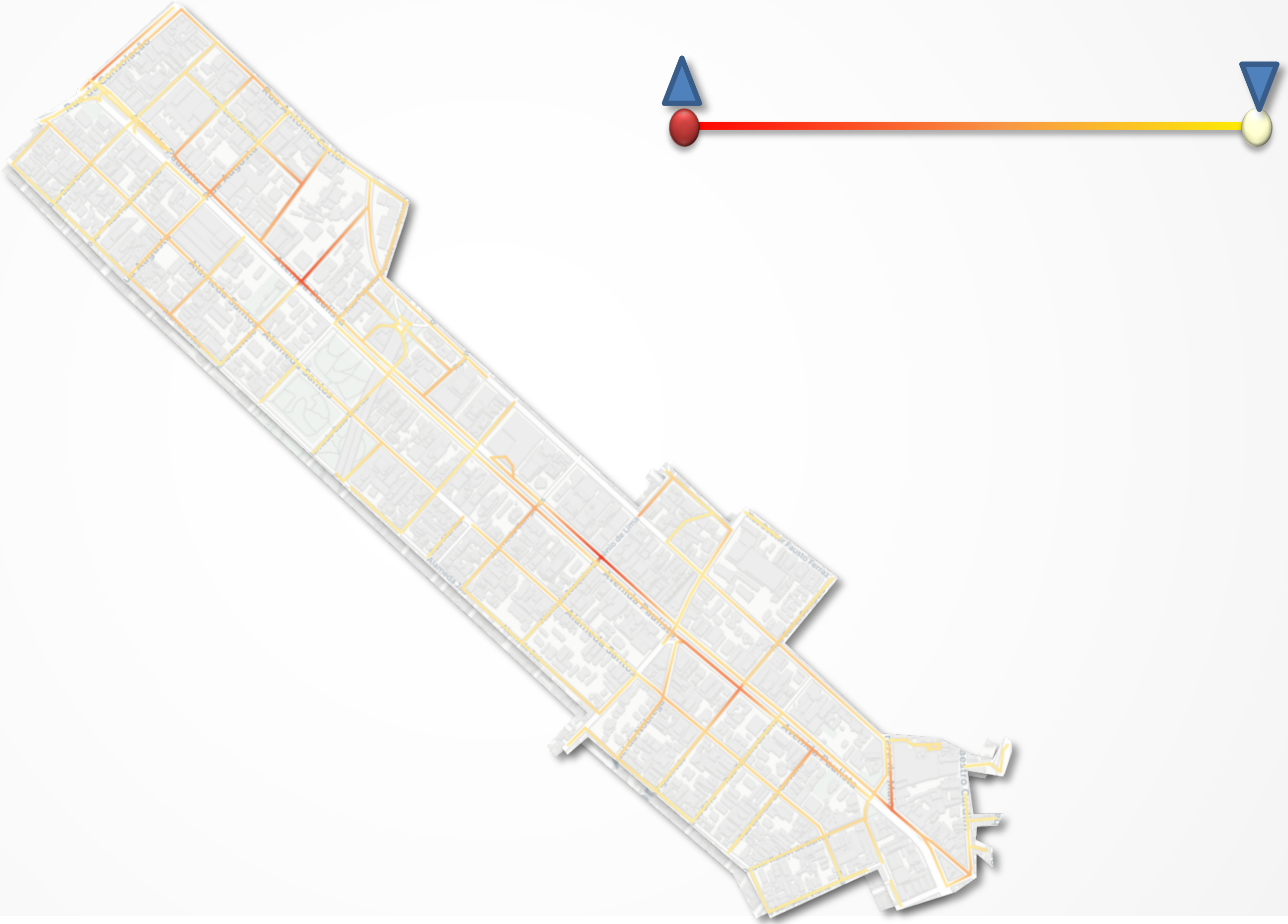

PARE
Google

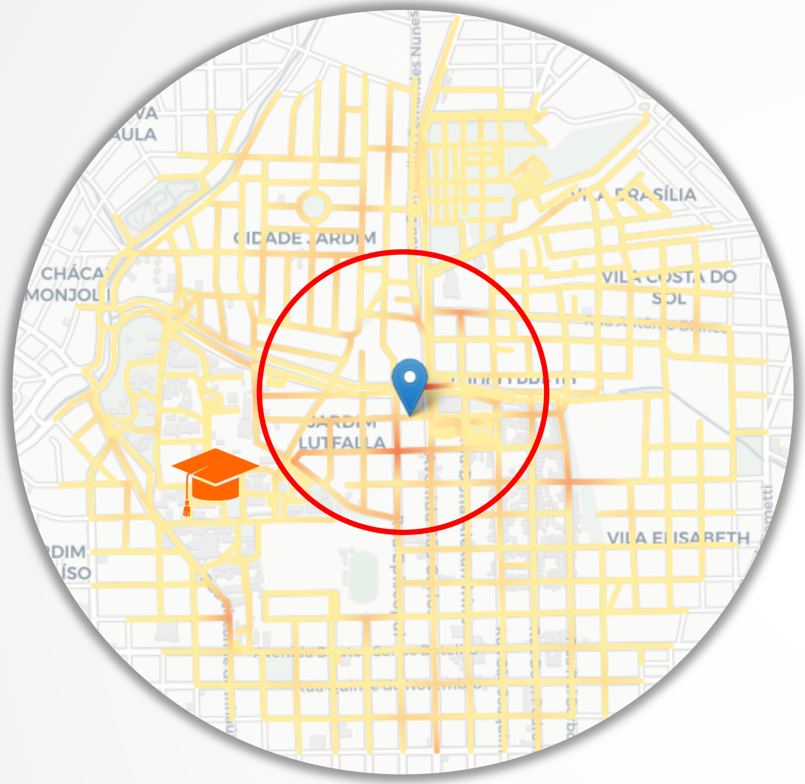
Results (3) - Regression

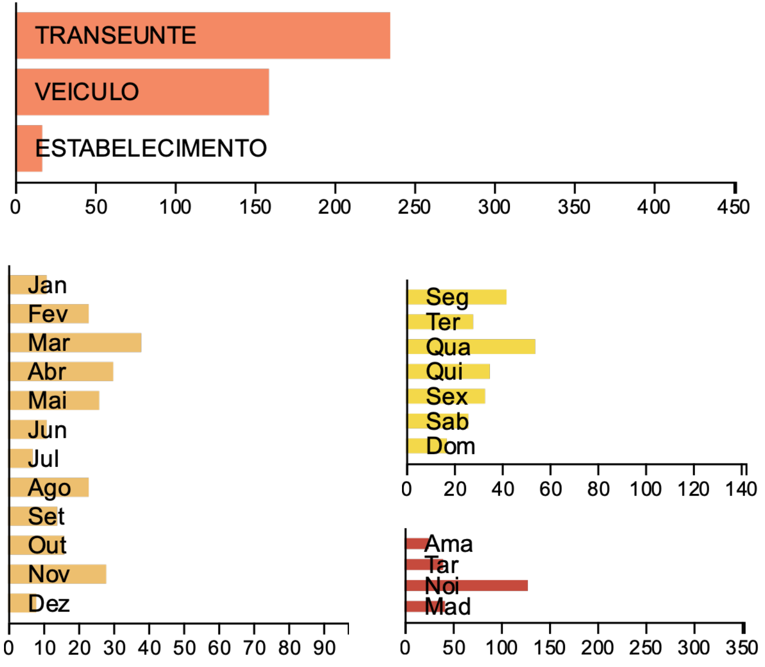
Ordinary Least Squares

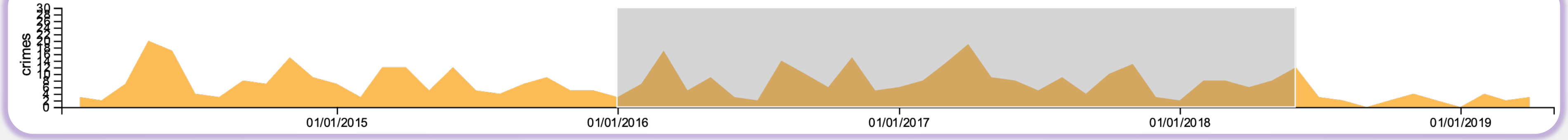
Infrastructure	Passerby R.	Establishment R.	Vehicle R.
BUS STOP	-	0.6	1.01
BAR	1.35	0.4	2.04

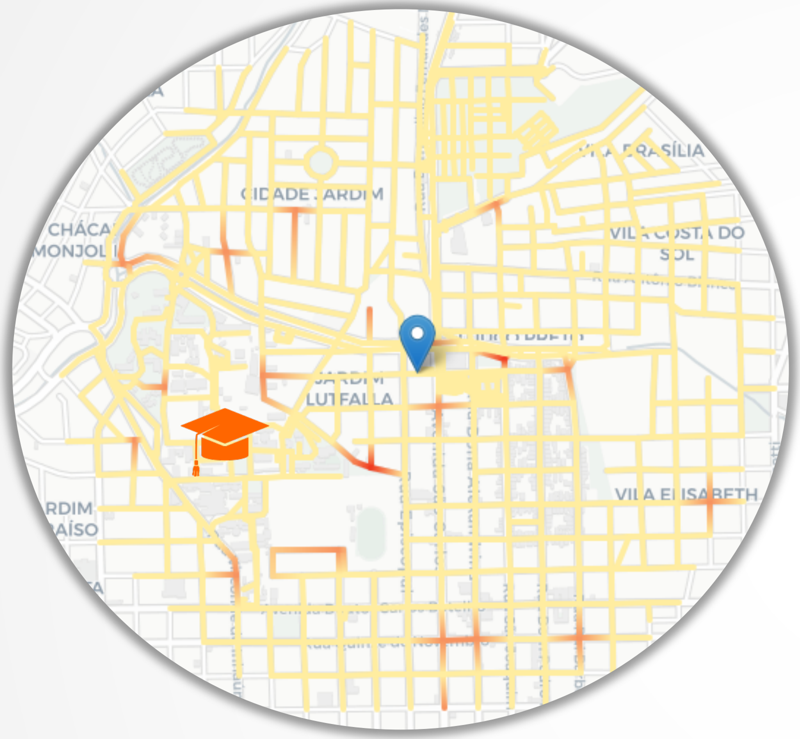
MIRANTE – Data Modeling

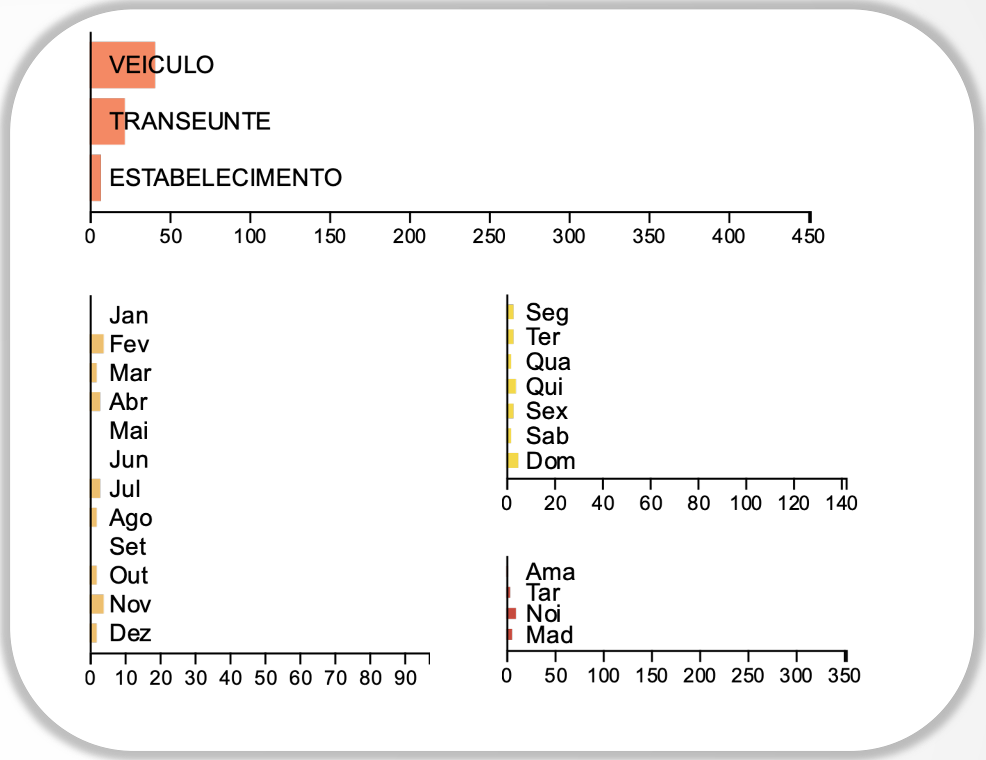




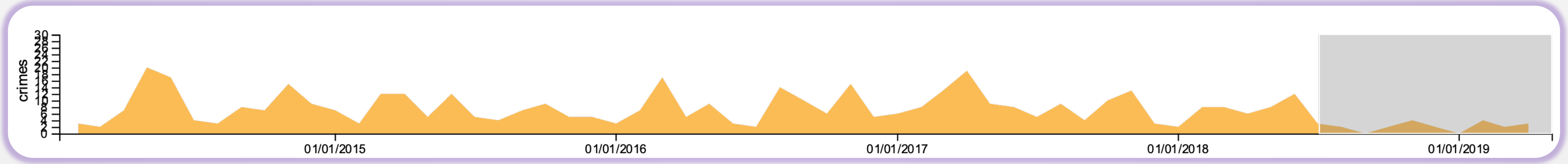

TRANSEUNTE
VEICULO
ESTABELECIMENTO
0 50 100 150 200 250 300 350 400 450
Jan
Fev
Mar
Abr
Mai
Jun
Jul
Ago
Set
Out
Nov
Dez
0 10 20 30 40 50 60 70 80 90
Seg
Ter
Qua
Qui
Sex
Sab
Dom
0 20 40 60 80 100 120 140
Ama
Tar
Noi
Mad
0 50 100 150 200 250 300 350


crimes
01/01/2015
01/01/2016
01/01/2017
01/01/2018
01/01/2019

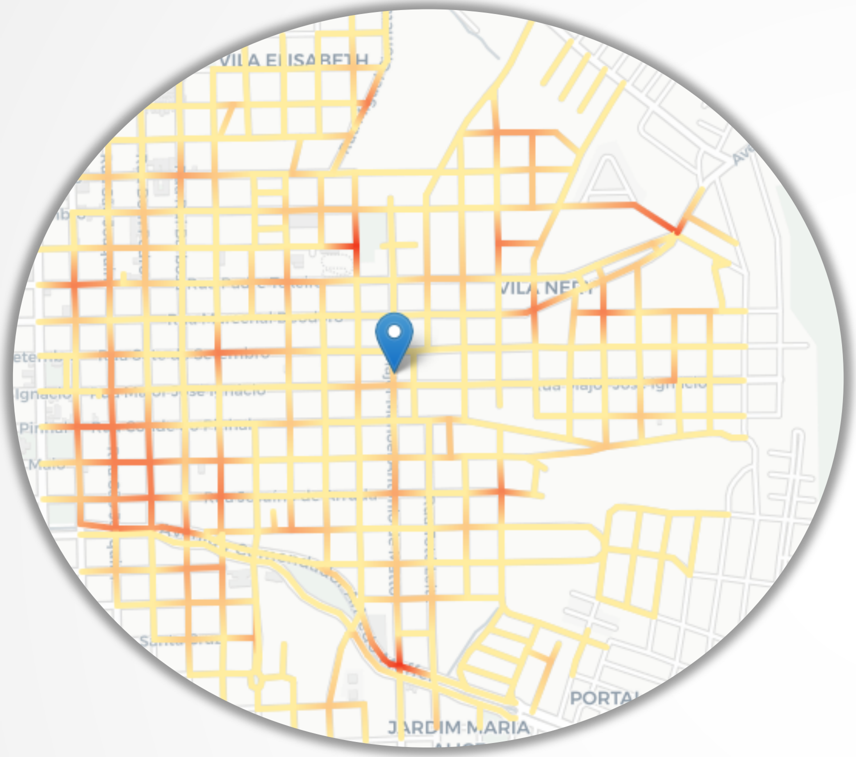


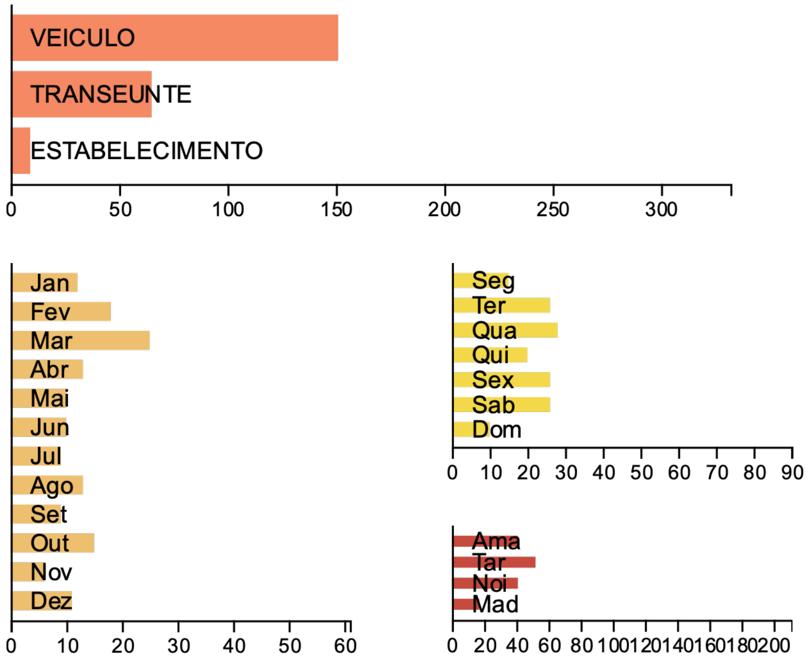


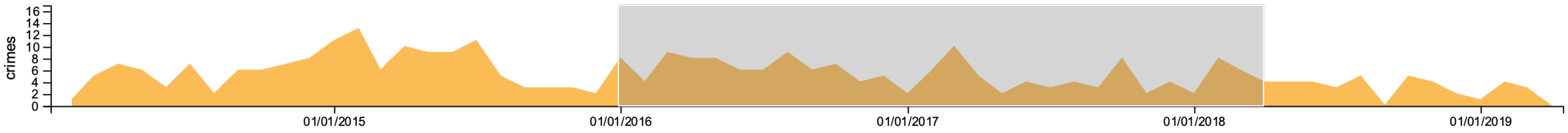
VEICULO
TRANSEUNTE
ESTABELECIMENTO
0 50 100 150 200 250 300 350 400 450
Jan
Fev
Mar
Abr
Mai
Jun
Jul
Ago
Set
Out
Nov
Dez
0 10 20 30 40 50 60 70 80 90
Seg
Ter
Qua
Qui
Sex
Sab
Dom
0 20 40 60 80 100 120 140
Ama
Tar
Noi
Mad
0 50 100 150 200 250 300 350

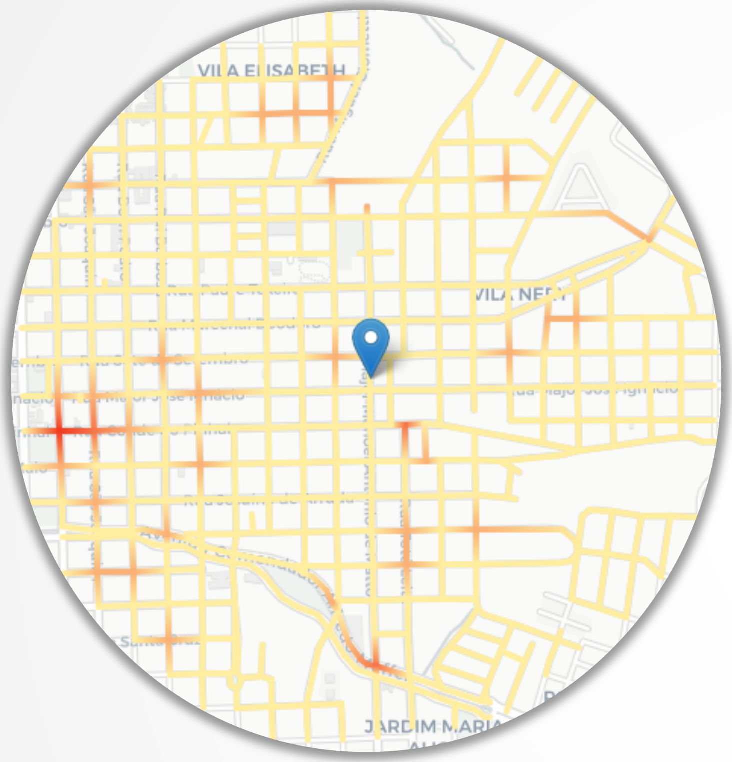


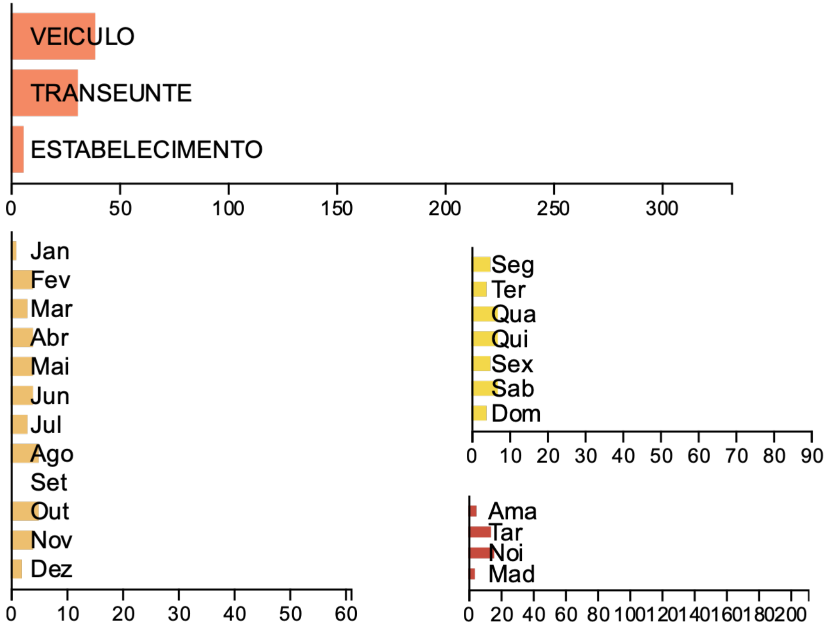
crimes
01/01/2015
01/01/2016
01/01/2017
01/01/2018
01/01/2019

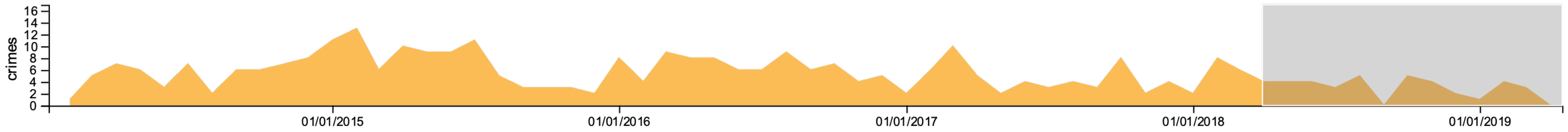

JARDIM MARIA


VEICULO
TRANSEUNTE
ESTABELECIMENTO
0 50 100 150 200 250 300
Jan
Fev
Mar
Abr
Mai
Jun
Jul
Ago
Set
Out
Nov
Dez
0 10 20 30 40 50 60
Seg
Ter
Qua
Qui
Sex
Sab
Dom
0 10 20 30 40 50 60 70 80 90
Ama
Tar
Noi
Mad
0 20 40 60 80 100 120 140 160 180 200


crimes
0 2 4 6 8 10 12 14 16
01/01/2015
01/01/2016
01/01/2017
01/01/2018
01/01/2019


VILA ELISABETH
JARDIM MARIA


VEICULO
TRANSEUNTE
ESTABELECIMENTO
0 50 100 150 200 250 300
Jan
Fev
Mar
Abr
Mai
Jun
Jul
Ago
Set
Out
Nov
Dez
0 10 20 30 40 50 60
Seg
Ter
Qua
Qui
Sex
Sab
Dom
0 10 20 30 40 50 60 70 80 90
Ama
Tar
Noi
Mad
0 20 40 60 80 100 120 140 160 180 200


crimes
16 14 12 10 8 6 4 2 0
01/01/2015
01/01/2016
01/01/2017
01/01/2018
01/01/2019

Conclusions:

- There is a direct relationship between the number of infrastructure and the number of crimes.
- The criminality directly affects the students’ performance, especially teenagers.
- We figured out some interesting criminal patterns on avenues and roads.
- Tensor decomposition can be applied with different data sources and different contexts.

Future Directions:

- We want to validate if, in fact, there is a relation between criminality and student's performance.
- We are going to do a study about homicides surroundings in São Paulo.
- We are going to implement some algorithms to find the best rank of the decomposition Tucker.

Thanks to CAPES for financial support

Thank you !


Germain García Zanabria
germaingarcia@usp.br


Luis Gustavo Nonato
gnonato@icmc.usp.br


Jaqueline Silveira
alva.jaque@usp.br


Afonso Paiva
apneto@icmc.usp.br


Sergio Adorno
marsadorno@usp.br


Marcelo Nery
mbnery@gmail.com