



ENGENYERIA INFORMÀTICA  
TREBALL FI DE GRAU

---

**Anàlisis dels accidents de tràfic  
a la ciutat de Barcelona  
(2010-2015)**

---

*Autor:*  
Nicolás Martín FORTEZA OCAÑA

*Tutor:*  
Dr. Jordi VITRIÀ

26 de junio de 2016

**resumen**

**resume**

# Índice

<b>1. Introducció</b>	<b>1</b>
<b>2. Com ens ho plantejarm</b>	<b>2</b>
2.1. Metodologies . . . . .	2
2.2. Software . . . . .	2
<b>3. Exploració de les dades</b>	<b>3</b>
3.1. Presentació dels DataSets . . . . .	3
3.2. Neteja de les dades . . . . .	4
3.3. Anàlisis inicials de les dades . . . . .	5
3.4. Conclusió de l'exploració . . . . .	10
<b>4. Anàlisi exhaustiu</b>	<b>11</b>
4.1. Plantejament de les hipotesis . . . . .	11
4.2. Enrequiment de lesDades . . . . .	11
4.3. Anàlisis . . . . .	11
4.4. Presentació de resultats . . . . .	11
<b>5. Conclusió i treball futur</b>	<b>12</b>
5.1. Conclusió . . . . .	12
5.2. Treball futur . . . . .	12
<b>6. Bibliografia</b>	<b>13</b>
<b>7. Anexos</b>	<b>14</b>

## 1. Introducció

En l'actualitat, el camp de la ciència de dades és un dels que més ha crescut en el món de l'enginyeria informàtica. Això és pel fet que les empreses i els governs han vist el potencial que tenen les dades, ja que es poden arribar a conclusions reveladores i crear patrons per a poder predir casos futurs. Aquests resultats poden fer que canviï el nostre punt de vista sobre segons quins temes o que, per exemple, una empresa pugui guanyar gran quantitat de beneficis en optimització.

Aquest treball es fa ressò d'aquest esclat del Data Science, però, en aquest cas, està enfocat als accidents de trànsit. En concret a la ciutat de Barcelona, una mostra prou gran per a treure conclusions de pes respecte als accidents de trànsit urbans.

El cos del treball se centra en 2 grans blocs. El primer bloc serà un aproximament inicial a les dades on les netejarem i farem una primera anàlisi de les dades on plantejarem petites hipòtesis amb les dades que tenim i farem algunes visualitzacions per així poder treure unes conclusions generals de les dades. Les conclusions extretes d'aquest primer bloc ens ajudaran a plantejar el bloc següent.

Aquest segon bloc és on plantejarem hipòtesis més complexes i que potser requereixen més dades de les que en un principi tenim. Per això, n'haurem de fer una cerca de noves dades per enriquir les que ja tenim o fer càlculs per generar-les a partir de les que ja tenim. Una vegada tinguem els nous datasets ja podrem passar a afirmar o refutar les hipòtesis plantejades.

L'estructura plantejada es basa en el mètode empíric-analític proposat per Neil J. Salkind (Cicle de la Investigació Científica).

## **2. Comens ho plantejarm**

### **2.1. Metodologies**

### **2.2. Software**

Utilitzar notebook  
utilitzar python i llibreries  
altres: leaflet, openrefine

## 3. Exploració de les dades

### 3.1. Presentació dels DataSets

Els datasets en els quals es basa aquest treball són els *Accidents gestionats per la Guàrdia Urbana a la ciutat de Barcelona*. Aquest datasets descriuen tots els accidents registrats per la Guàrdia Urbana a la ciutat de Barcelona des del 2010 fins al 2015, inclosos. Estan accessibles públicament a la plana web Open Dat BCN, pàgina de l'Ajuntament de Barcelona on posa a l'abast del ciutadà dades dels organismes municipals.

Per cada any hi han publicades 5 taules diferents:

- **Taula general:** En aquesta taula trobem les dades generals de cada accident. Comença amb les dades de temporals i de localització (aquestes dades són comunes a totes les taules), i prossegueix amb el nombre de persones implicades segons estat (ferits lleus, ferits greus o morts).
- **Taula de causalitat:** En aquesta taula hi ha la part comuna (localització i moment) hi ha una altra columna que descriu les causes mediatas. Aquesta columna no descriu les causes que es van trobar posteriorment en les investigacions, solament inclou les causes que la Guàrdia Urbana va observar sobre el terreny.
- **Taula de les personnes implicades:** En aquesta taula a banda de la localització espacial i temporal, també trobem una descripció de les personnes implicades en l'accident. Aquí podem veure coses com en quin vehicle viatjava, de quin sexe és, quina edat tenia i com l'ha afectat (si està, per exemple, ferit o mort).
- **Taula de tipologia:** en aquesta taula hi ha una columna on ens descriuen com ha sigut l'accident: si ha sigut una caiguda, una col·lisió, una bolcada, etc.
- **Taula de vehicles implicats:** En aquesta taula a banda de la localització espacial i temporal, també trobem una descripció de cada vehicle implicat. Hi ha columnes que descriuen el tipus de vehicle, el model, la marca i el color, també podem saber quin tipus de carnet i amb quina antiguitat.

Per a l'anàlisi inicial solament utilitzarem la taula general, ja que ja d'aquí podem treure bones conclusions i idees de per on continuar la investigació. Així quan fem la segona part, omplirem la nostra taula general netejada amb les altres taules. Intentarem posar la màxima informació possible a la taula general per així evitar el màxim possible treballar amb més d'una taula a l'hora.

### 3.2. Neteja de les dades

Tota base de dades omplerta per humans pot tenir errors ortogràfics o de tipografia. Això fa que quan ho carreguis en un programa pugui entendre dos atributs iguals per diferents si estan escrits diferents. Per aquesta raó aquesta part és una de les més importants quan es treballen amb dades, ja que treballar amb unes dades sense refinjar pot portar a conclusions errònies.

El primer que hem hagut d'afrontar és el fet de les codificacions de les taules. Encara que sembli una ximpleria pot portar a molts errors, ja que pot carregar alguns dels caràcters com si fossin algun altre. En el nostre cas algunes de les taules estaven en **Latin-1(ISO 8859-1)** i les altres en **Europa occidental(CP850)**. Per trobar-hi el format corresponent va caldre una exhaustiva cerca, ja que estem acostumats al ja quasi estandarditzat **UTF-8**.

#### Refinació de les dades amb OpenRefine.

Seguidament vam passar a fer el refinatge de les columnes amb el OpenRefine.

OpenRefien té una opció per a cada columna (**cluster and edit...**) que ens permet comparar totes les paraules amb algorismes com el *fingerprint* o la *distància de Levenshtein*, ajuntar-les i editar-les totes a l'hora.

En el cas de les nostres dades hem corregit principalment la columna de carrers, ja que hi havia un parell d'errors amb accentuació i majúscules.

#### Correcció de capçaleres

Un problema amb la llibreria *Pandas* és que dóna errors al accedir-hi a capçaleres amb caràcters utf-8, així que una vegada carregades les taules al nostre projecte de *ipython notebook*, procedim a canviar-les.

Per a canviar-les hem fet un programa que, agafant totes les capçaleres, primer que posa tots el caràcter en minúscula i que tregui tots els accents. Posteriorment fa un split per espais i així pot posar en majúscula la primera lletra de cada paraula. Per finalitzar ho ajunta tot en una única paraula i treu els possibles caràcters estranys que hagin quedat.

### Canvi del sistema de coordenades

Una altra cosa que vam fer a l'inici del projecte, va ser canviar el sistema de coordenades de les taules. A totes les taules hi ha una columna que dóna les coordenades geogràfiques de l'accident, però l'únic inconvenient és que estan al sistema UTM(*Sistema de Coordenades Universal Transversal de Mercator*). Com per nosaltres és més fàcil treballar en LatLon( Latitud-Longitud), hem hagut de fer un petit programa que mapeja les columnes i crea una nova columna amb les coordenades en LatLon. Aquesta conversió l'hem fet gràcies a la llibreria *UTM* per a *Python*.

### 3.3. Anàlisis inicials de les dades

Per aquesta part hem creat uns quants programes que ens faciliten la feina.

Primer hem fet unes quantes funcions que ens ajuden a fer operacions amb els *DataFrames* de *Pandas*, com per exemple, contar elements o contra elements i posar-ho en diccionaris. Com quasi tots els resultats els hem treballat amb diccionaris, també tenim funcions per a diccionaris, per exemple, que ens dóna un top 10 d'elements d'un diccionari, en retorna una llista ordenada o fa la mitja dels seus valors.

Després tenim altres funcions que en permeten dibuixar amb la llibreria *matplotlib* solament passant-hi un diccionari. Aquestes funcions ens dibuixen diagrames de barres, diagrames *pie* i diagrames de barres apilats.

#### Accidents totals

Comencem fent una mirada general als accidents. En la figura 1 veiem l'evolució del total d'accidents per any. Podem observar un ascens dels accidents durant els últims dos anys. Aquests dos últims anys estan també per damunt de la mitja dels anys (9380 accidents a l'any).

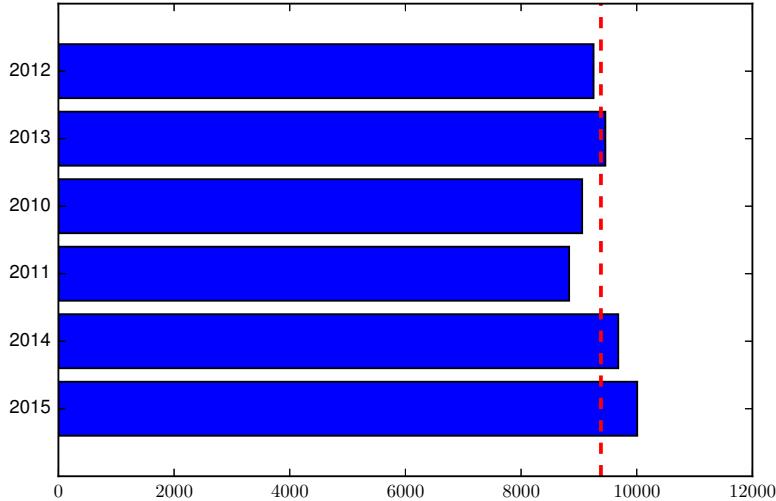


Figura 1: Total d'accidents per any

### Mapes de calor

Per fer-nos una idea dels accidents respecte a la seva localització, hem fet mapes de calor, així sabem en quines zones es concentren més accidents i en quines no.

En el mapa de 2010 (figura 7 pàgina 16) observem que es va concentrar molts accidents per la zona de carrer Aragó amb carrer Balmes, rambla de Catalunya i Passeig de Gràcia. En canvi, en mapes com el del 2011 (figura 8 pàgina 17) no hi ha tanta concentració en una zona, però hi ha més zones de concentracions d'accidents, com per exemple en aquest cas, hi ha concentracions en Gran Via de les Corts Catalanes, a l'altura d'Universitat i Comte Urgell.

En general sempre s'aprecia grans concentracions d'accidents a zona de l'Eixample, entre Diagonal i Gran Via, i entre carrer de Sants i Passeig de gràcia. Però en mapes com el de 2014 (figura 11 pàgina 20) i el de 2015 (figura 12 pàgina 21), també veiem concentracions notables d'accidents per la zona de Clot i de Bogatell.

## Accidents per districte

Els numero d'accidents per districte (figura 2 i figura 13 pàgina 22) ens fa veure com els d'accidents a l'Eixample (2710 de mitja anual) és el 28% dels accidents a tota Barcelona. Per darrere venen districtes com Sant Martí, Sarrià-Sant Gervasi i Sants-Montjuïc que contenen un 10% dels accidents. La resta de districtes ronden el 5%.

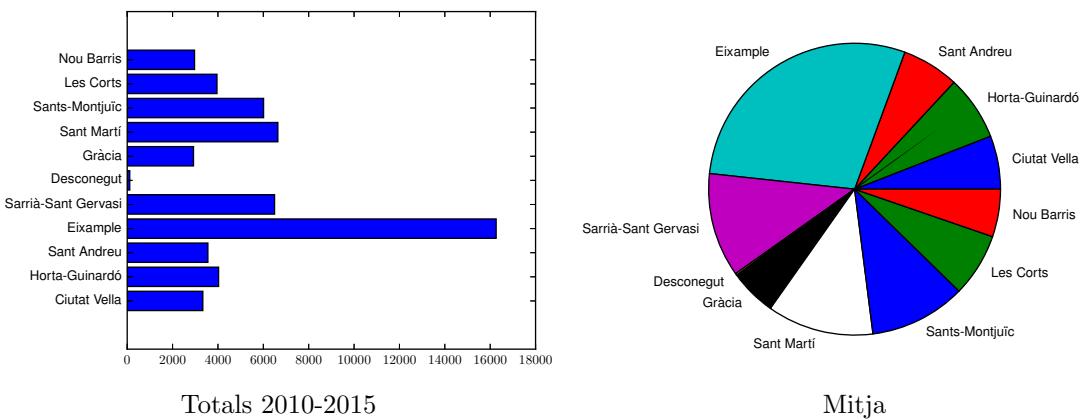


Figura 2: Accidents per districte

## Accidents per barri

Respecte als accidents, solament hem mirat els 10 amb més accidents (figura 3 i figura 14 pàgina 23), ja que hi ha 73 barris a Barcelona. El barri més destacat és la Dreta de l'Eixample, que té una mitja de 1060 accidents per any, 554 més que el segon, l'Antiga Esquerra de l'Eixample, que té una mitja de 506 accidents per any. Barris per sobre del 300 accidents per any també trobem Sant Gervasi – Galvany (amb 422 accidents per any), la Nova Esquerra de l'Eixample (amb 353 accidents per any) i la Sagrada Família (amb 301 accidents per any). Entre 300 i 200 accidents trobem les Corts, el Fort Pienc, el Poble Sec, Sant Antoni i la Maternitat i Sant Ramon.

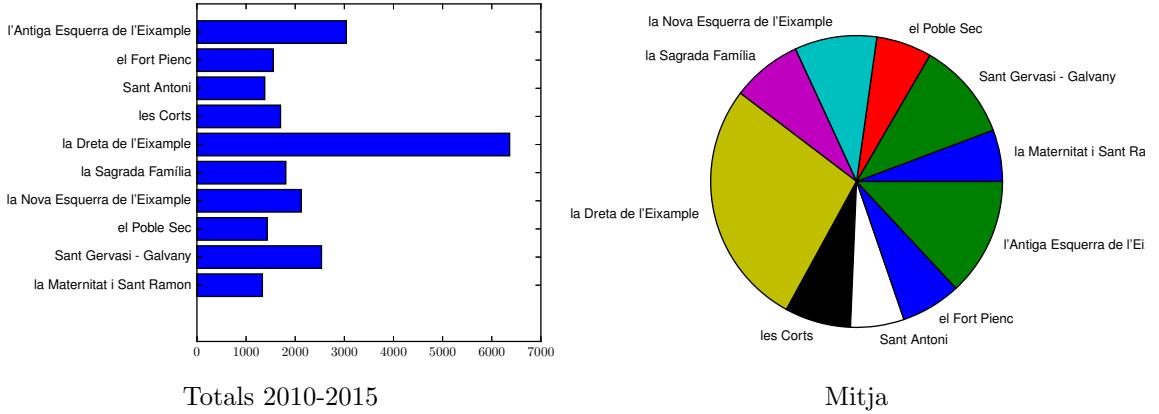


Figura 3: Accidents per barri

### Accidents per carrer

En el rànquing d'accidents per carrers (figura 4 i figura 15 pàgina 24) veiem que els dos primers són, amb diferència, la Gran Via de les Corts Catalanes, amb una mitja de 471 accidents per any, i Diagonal, amb una mitja de 402 accidents per any. Posteriorment ve Aragó amb 227 accidents l'any. La resta de carrers ja tenen una mitja d'accidents inferior a 200, com per exemple, Litoral (Llobregat) amb 172 o Meridiana amb 166.

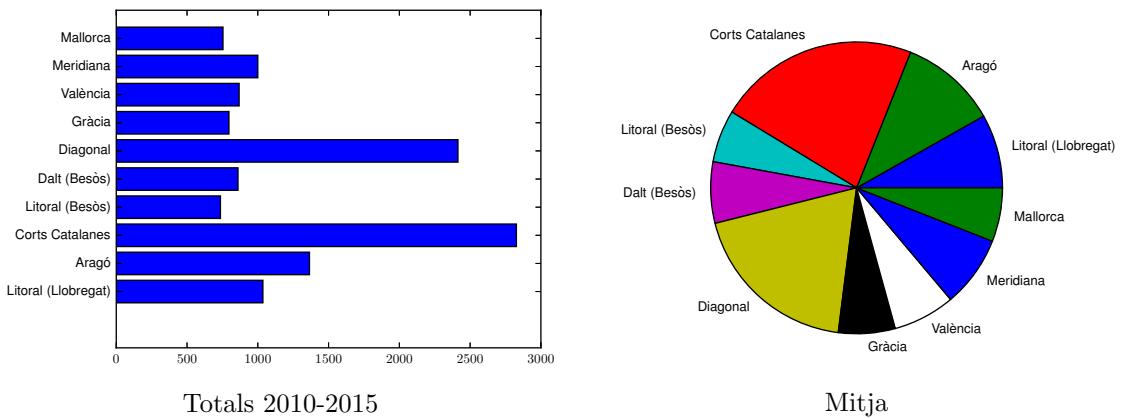


Figura 4: Accidents per carrer

## Accidents per mes

Quan mirem els accidents per mes (figura 5 i figura 16 pàgina 25) de l'any, veiem que sembla que a la tardor hi ha més accidents, ja que l'Octubre i el Novembre són els dos mesos amb més accidents (amb 837 i 835 accidents de mitja respectivament). Entre el mes amb més accidents (Octubre amb 837) i el penúltim (Gener amb 743) hi ha solament una diferència de quasi 100 accidents, en canvi, la diferència entre el penúltim i Agost (l'últim amb 576) hi ha una diferència de 167 accidents.

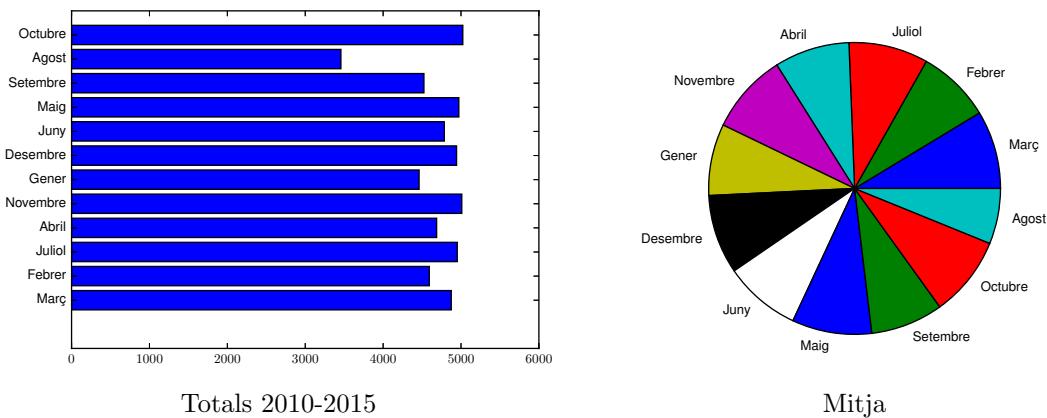


Figura 5: Accidents per mes

## Accidents per dia de la setmana

Estudiant els accidents pels dies de la setmana (figura 6 i figura 17 pàgina 26), veiem que el dia de la setmana amb més accidents de trànsit és el di-vendres amb una mitja de 1632. En canvi el dia de la setmana laboral amb menys accidents el dilluns amb 1423. De mitja durant la setmana laboral hi ha una mitja 1519 accidents, però durant el cap de setmana baixen els números. El dissabte amb una mitja de 1010 i el diumenge (el dia de la setmana amb menys accidents) amb 771 accidents de mitja.

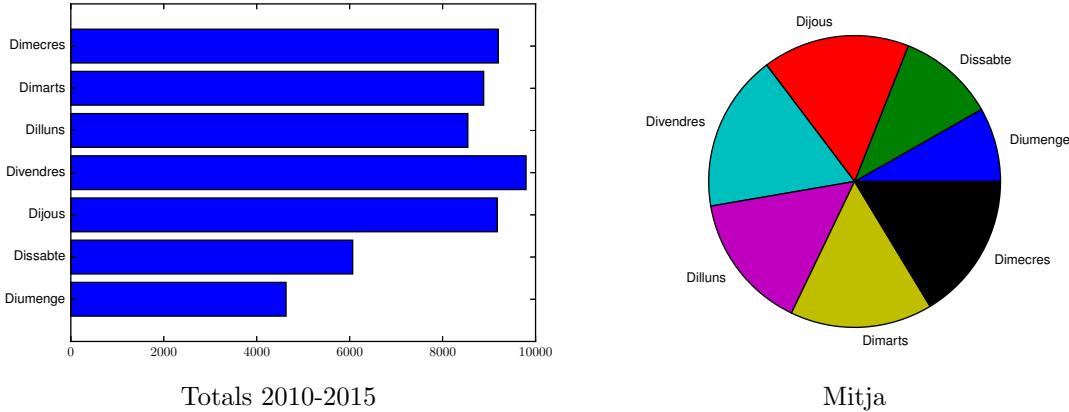


Figura 6: Accidents per dia de la setmana

## Accidents per TODO

### 3.4. Conclusió de l'exploració

Després de l'estudi geogràfic que hem fet, podem corroborar part del que hem vist inicialment al mapa. Com ben havíem avançat al mapa de calor, els accidents principalment es concentren en la zona de l'eixample, dividits entre Aragó i Gran via de les Corts Catalanes, dos dels carrers més transitats de Barcelona. A banda d'aquestes també trobem el Carrer Diagonal, que és justificable el seu nombre d'accidents, ja que és l'arteria principal del tràfic de la ciutat.

En qüestió temporal ens atrevim a afirmar que els dies no lectius i més els dies laborals, fan que descendeixi els accidents. El cap de setmanes descedeixen molt els accidents com també mesos com Agost. Per algun fet, també hi ha més accidents el divendres que els dilluns, això podria estar degut a l'actitud de la gent en aquests dos dies. Un altre punt curiós és que a la tardor hi ha més accidents, sobretot a l'Octubre i al Novembre.

## **4. Anàlisi exhaustiu**

- 4.1. Plantejament de les hipotesis**
- 4.2. Enrequiment de les dades**
- 4.3. Anàlisis**
- 4.4. Presentació de resultats**

## **5. Conclusió i treball futur**

### **5.1. Conclusió**

### **5.2. Treball futur**

## 6. Bibliografia

## **7. Anexos**

### **Mapes**

## Plots

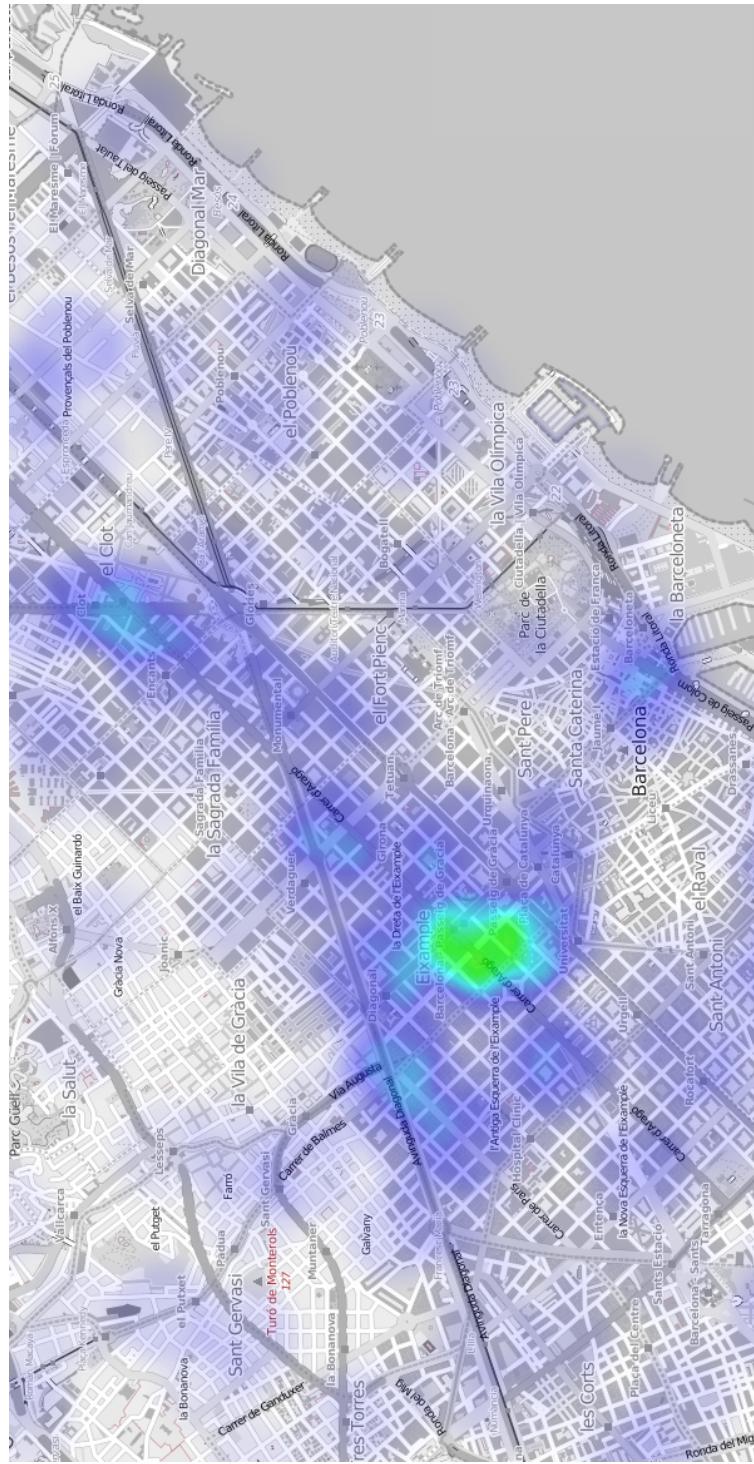


Figura 7: Mapa de calor 2010



Figura 8: Mapa de calor 2011

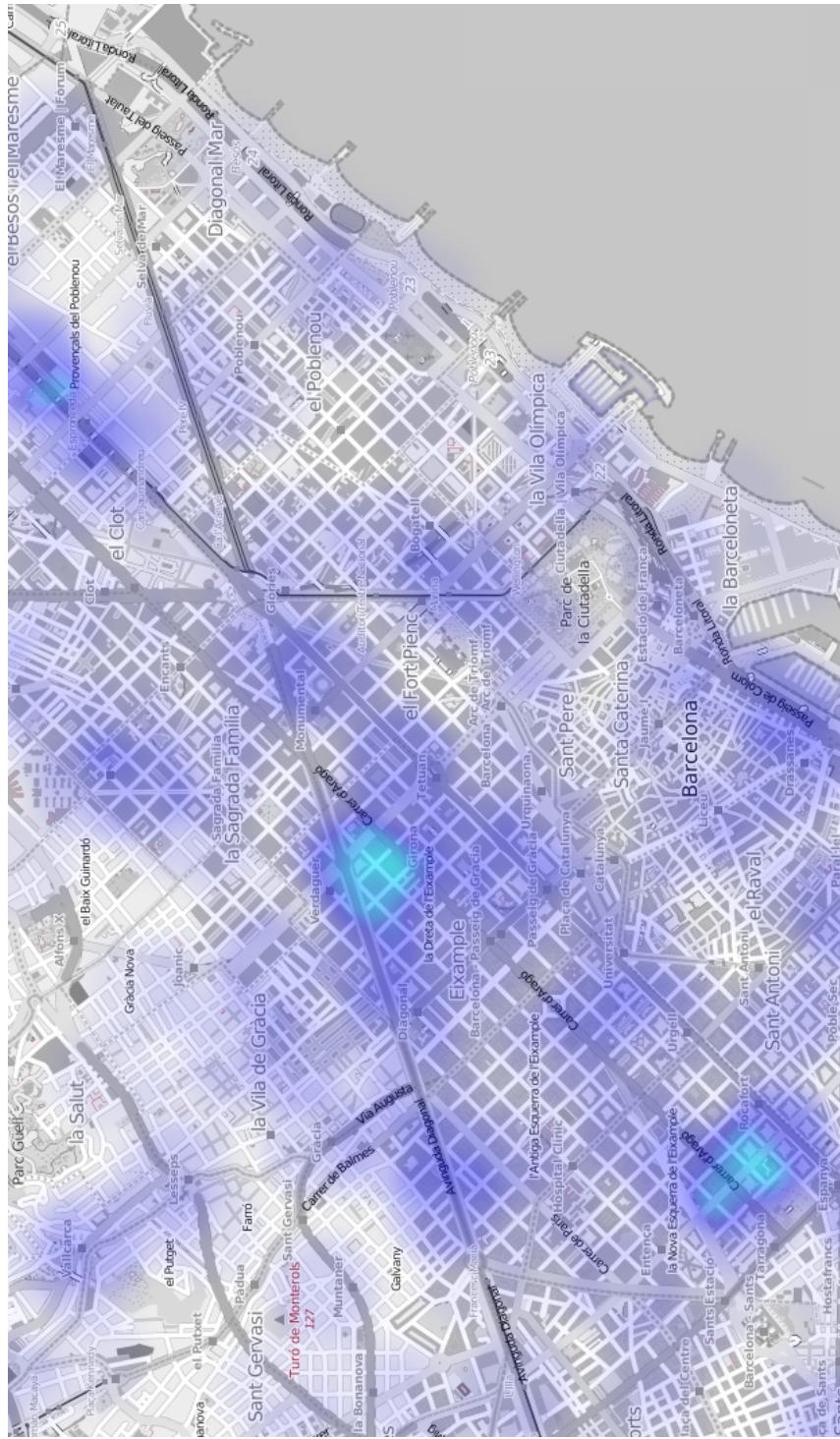


Figura 9: Mapa de calor 2012

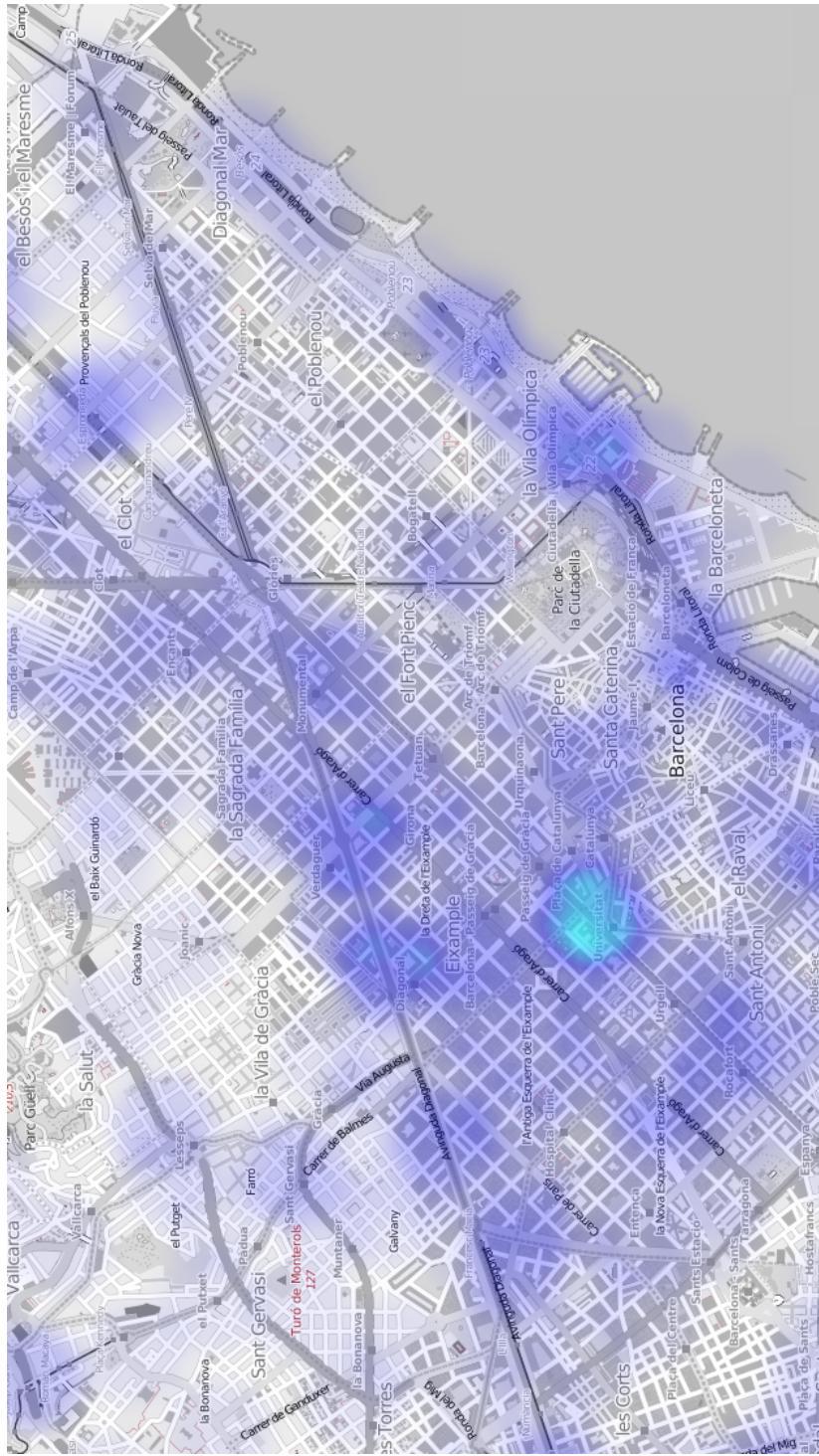


Figura 10: Mapa de calor 2013

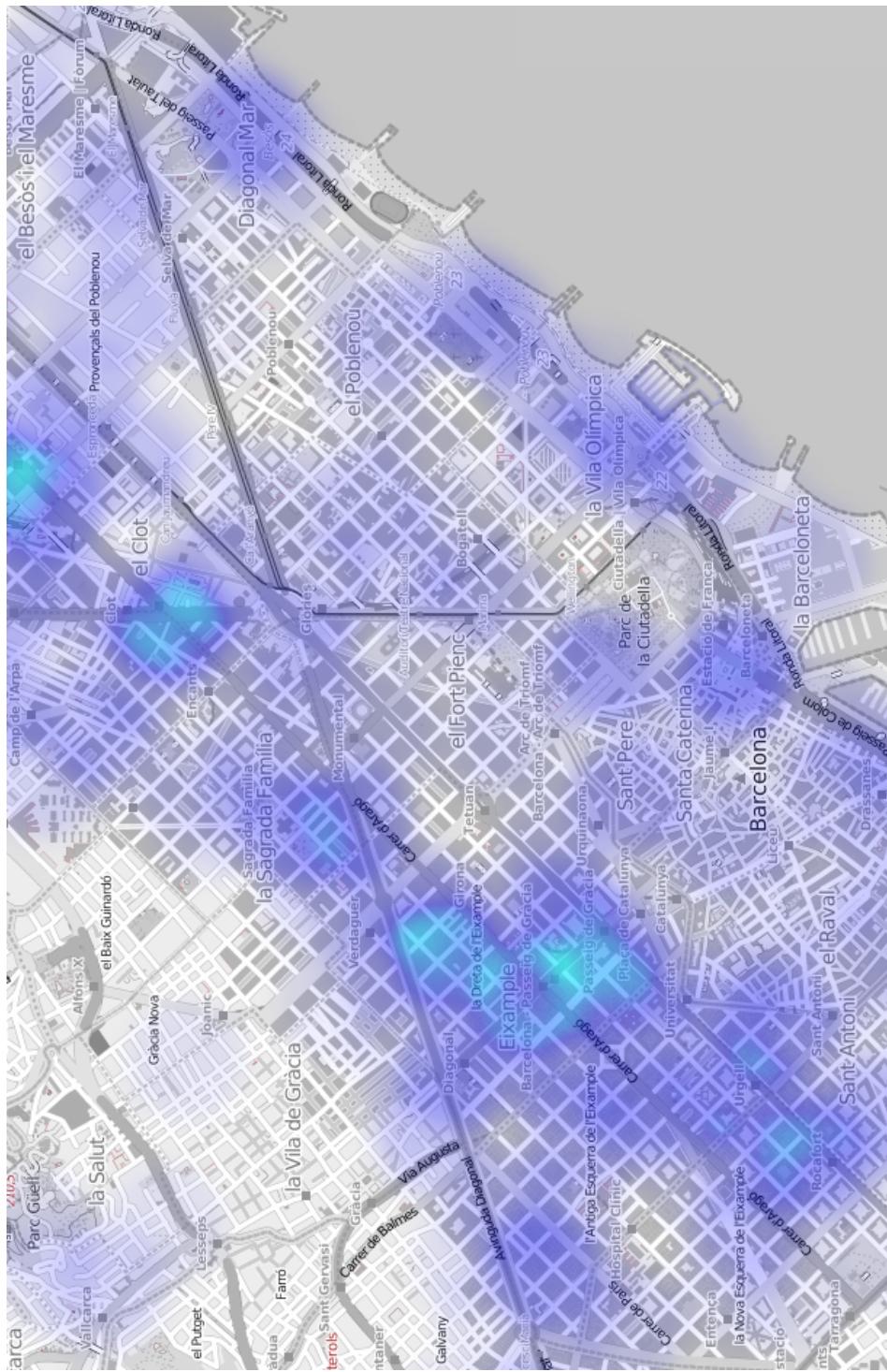


Figura 11: Mapa de calor 2014

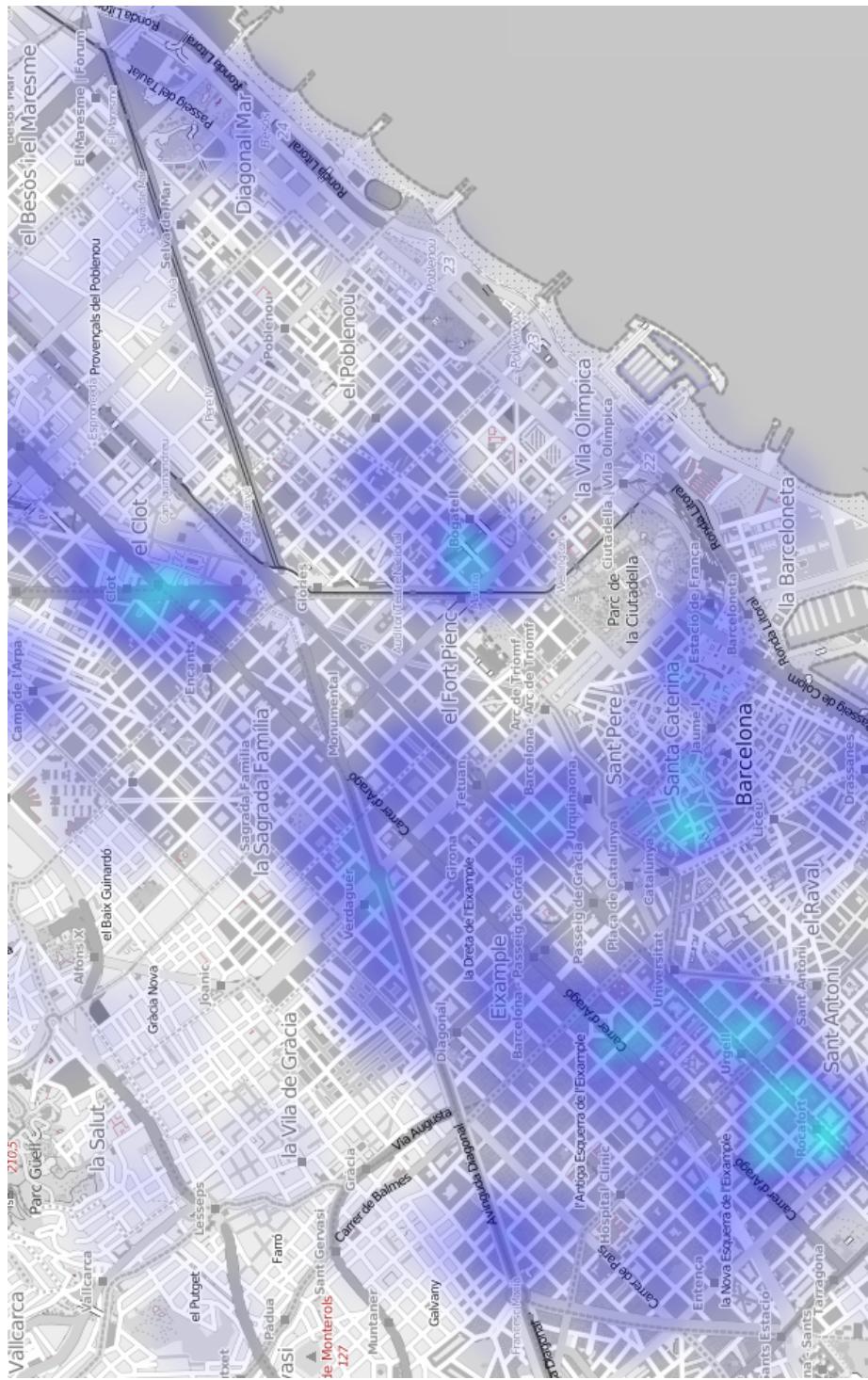


Figura 12: Mapa de calor 2015

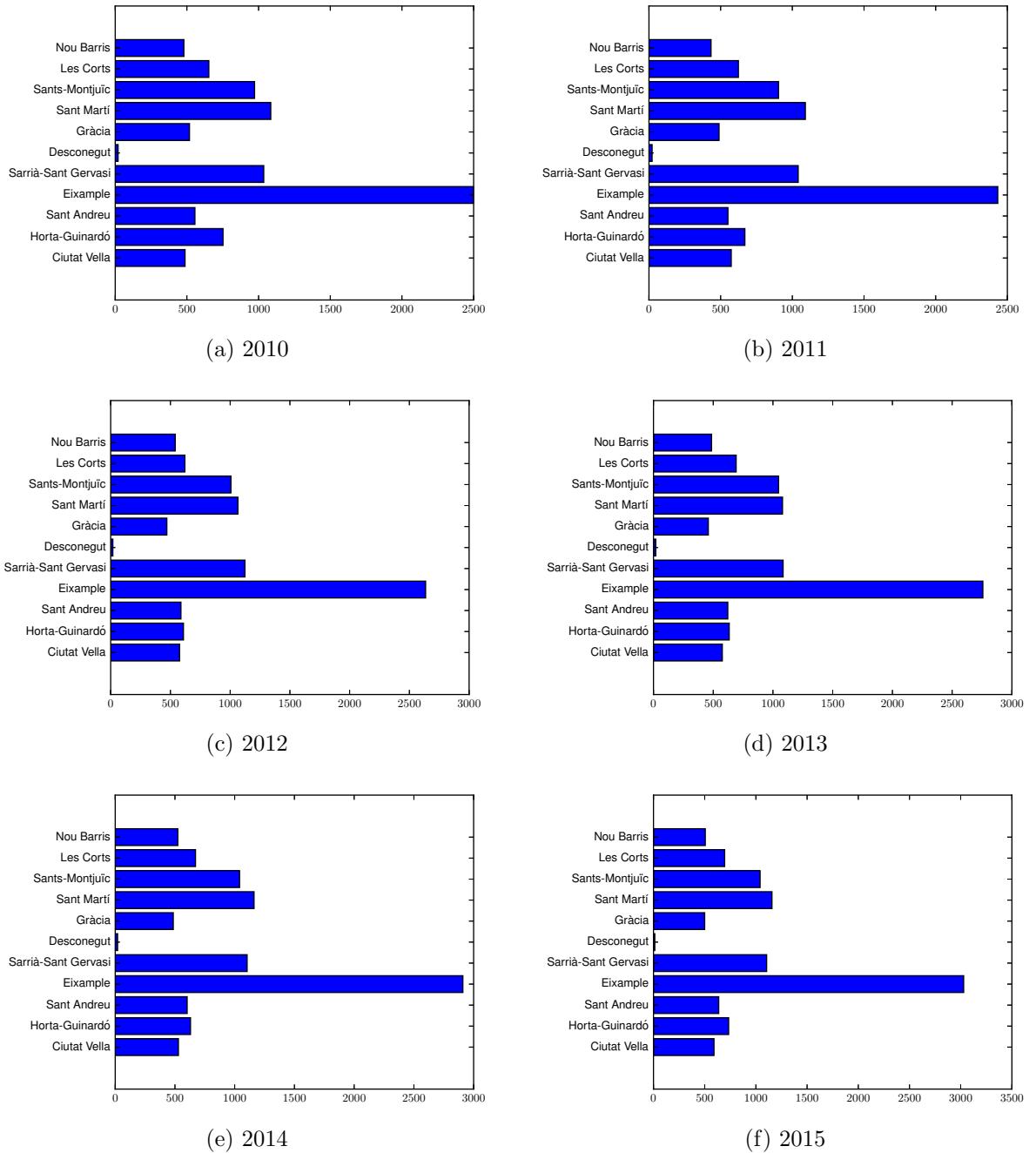


Figura 13: Accidents per Districtes

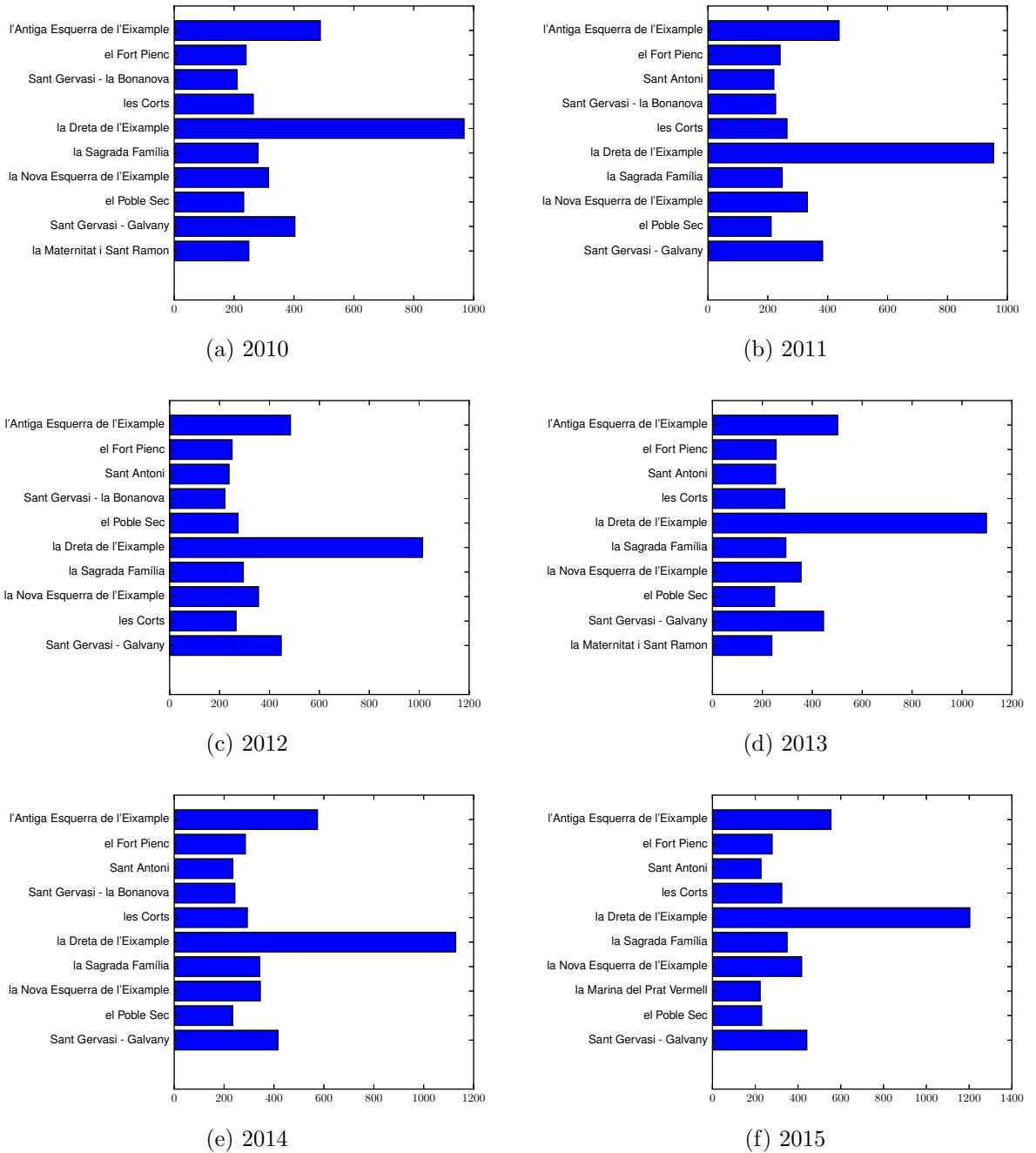


Figura 14: Accidents per Barris

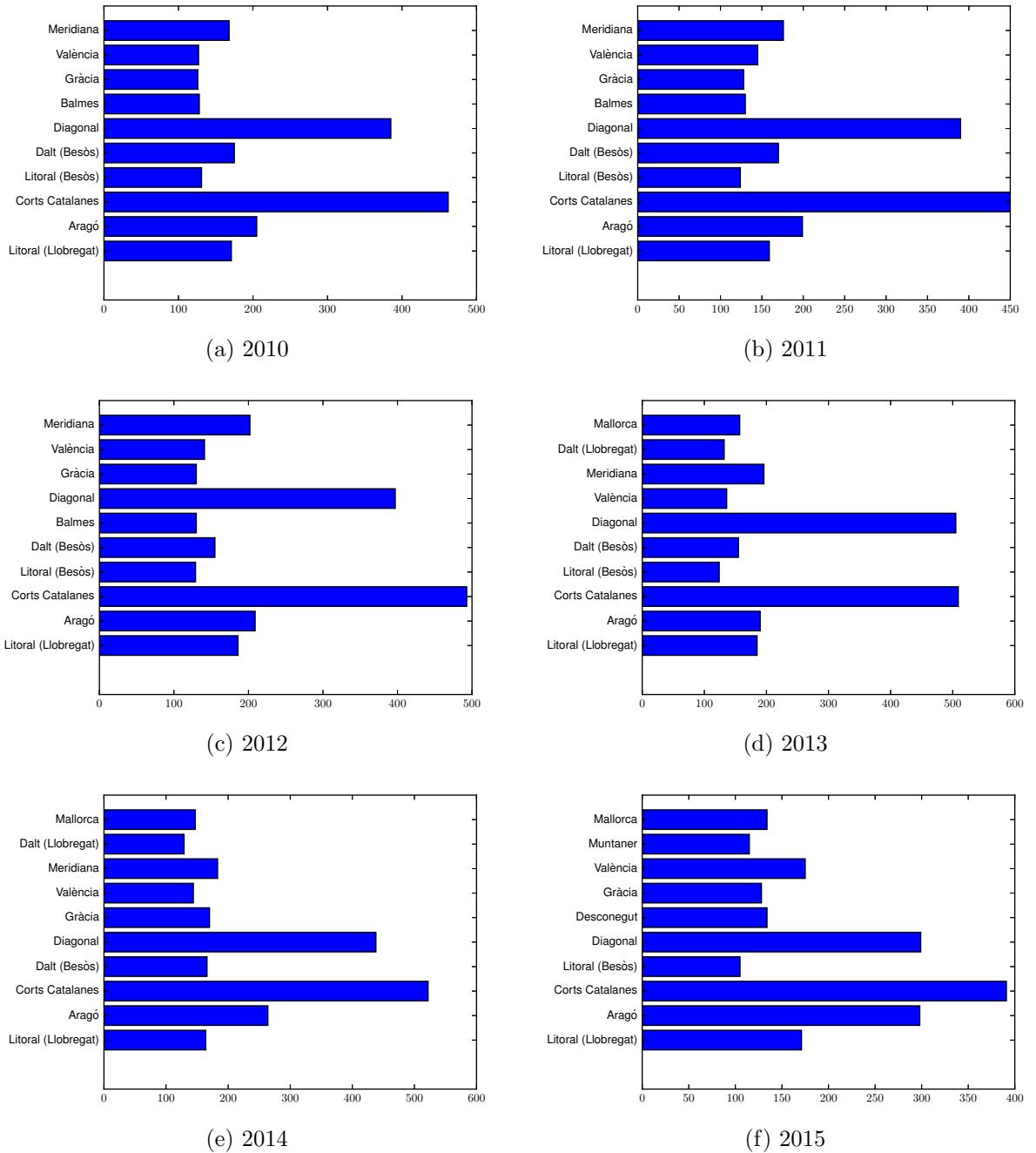
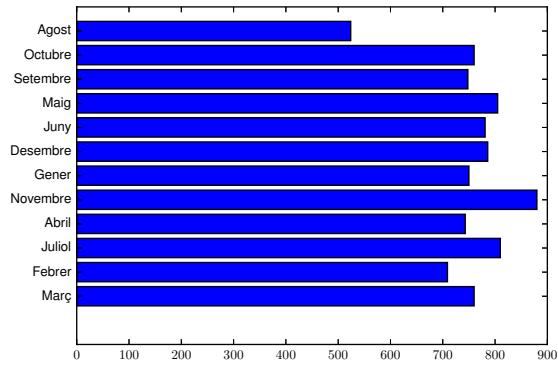
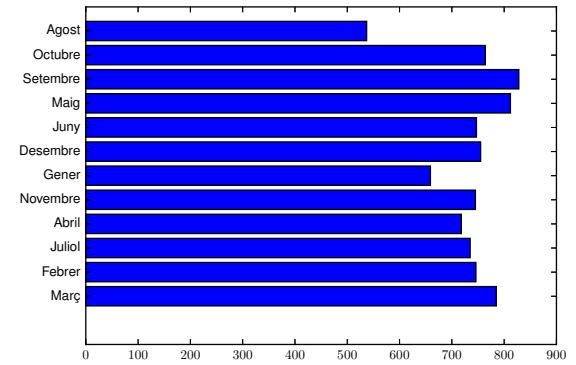


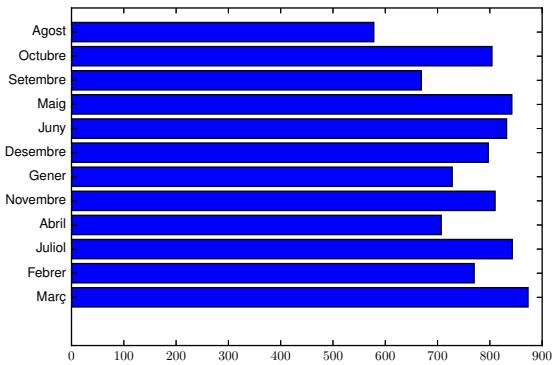
Figura 15: Accidents per Carrers



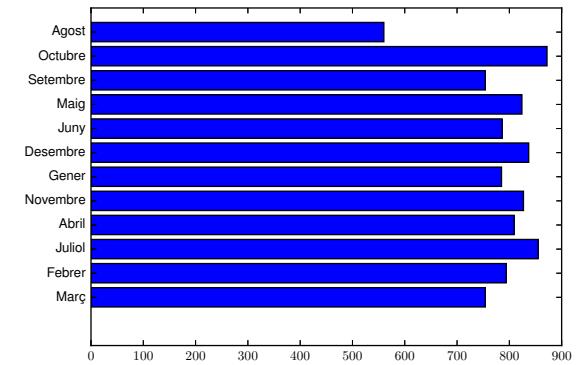
(a) 2010



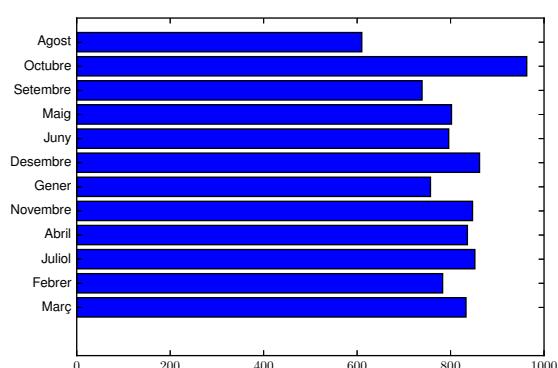
(b) 2011



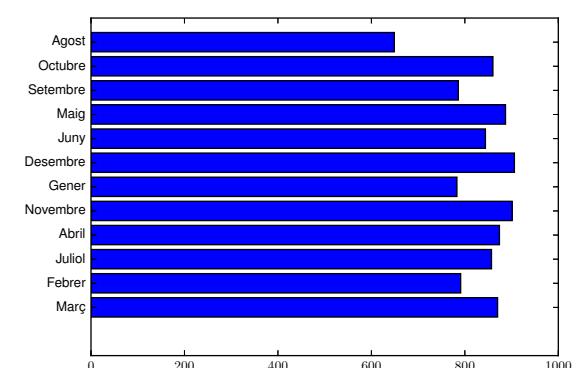
(c) 2012



(d) 2013

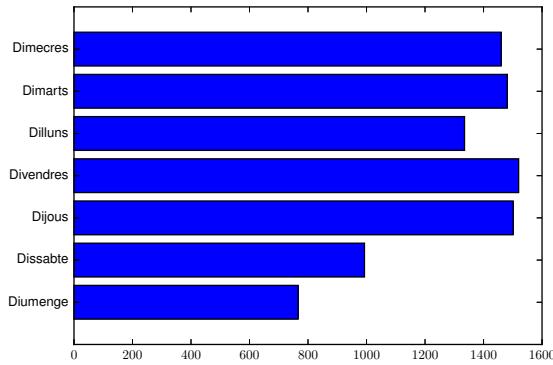


(e) 2014

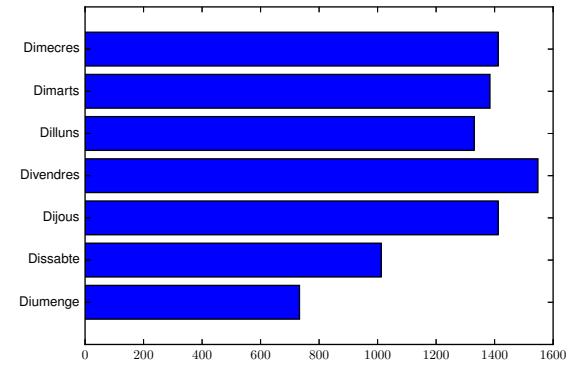


(f) 2015

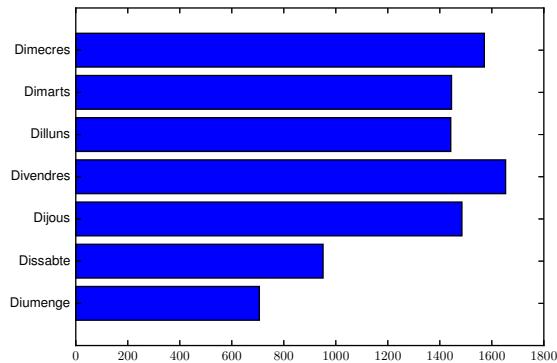
Figura 16: Accidents per mes



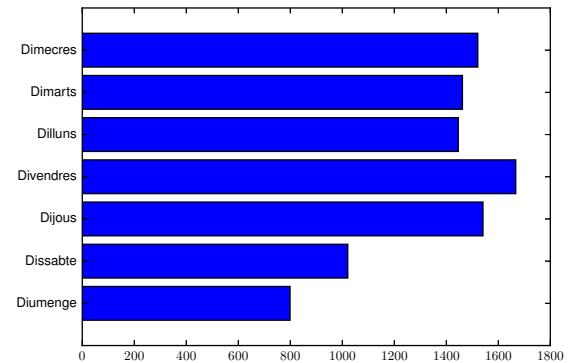
(a) 2010



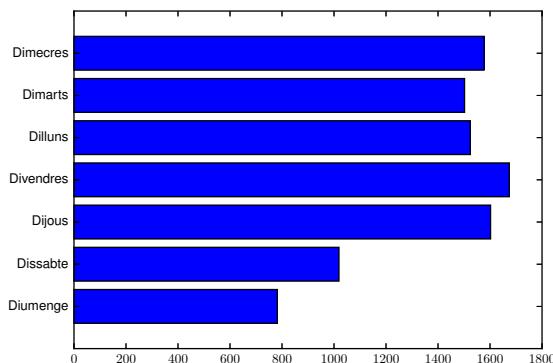
(b) 2011



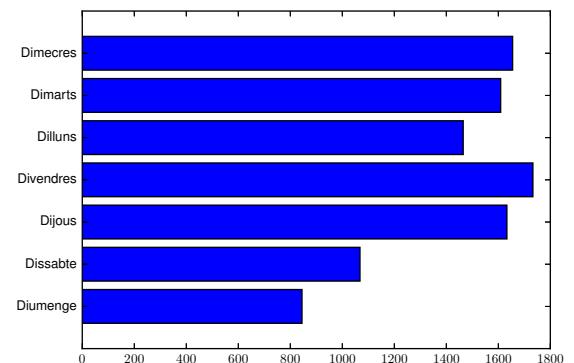
(c) 2012



(d) 2013



(e) 2014



(f) 2015

Figura 17: Accidents per Setmanes