



 #DataScienceEnPython

*Meetup*

**DATA SCIENCE EN  
PYTHON**

**MIÉRCOLES 7 DE MARZO**



/THECUBEMadrid



THECUBE Madrid



@THECUBEMadrid



/thecubemadrid

Sponsored by:





**Open Session**  
en TheCUBE Madrid

**15<sup>th</sup> March**  
**18:30**



**Data Science** Internet Things of



# Unlimateck Founders

Entrepreneurs in Residence at Thecube



#DataScienceEnPython



# Thecube=(Tech+Innovation+Education)

Inside Unlimiteck co-creation model, innovation will occur by design not by exception.



**The home of tech startups in Madrid**



**IoT Education facilities for professionals and graduates**



**Drive co-innovation with Partners and Startups**



**#DataScienceEnPython**

[www.unlimiteck.com](http://www.unlimiteck.com)

© 2018 unlimiteck Company Builder. All Rights Reserved.

# Agenda

**19:00** Bienvenida y objetivos del grupo DataScience  
Python TheCUBE Madrid. **Carlos Picazo**

**19:15** Clasificación de cáncer de mama utilizando sklearn  
**Diego García Morate**

**19:40** Una aplicación práctica del data science: Prevención  
de incendios forestales. **Ernesto Padilla Martínez, David  
Márquez e Ignacio Rivera.**

**20:05** Mesa Redonda Q&A

**20:20** Networking: Refrescos y cervezas

**21:00** Fin de evento



#DataScienceEnPython





#DataScienceEnPython

Objetivos del grupo de *Meetup*

Data Science Python Madrid TheCube



```
print("Hello, world!")
```







#DataScienceEnPython

# ¿Qué es Data Science?

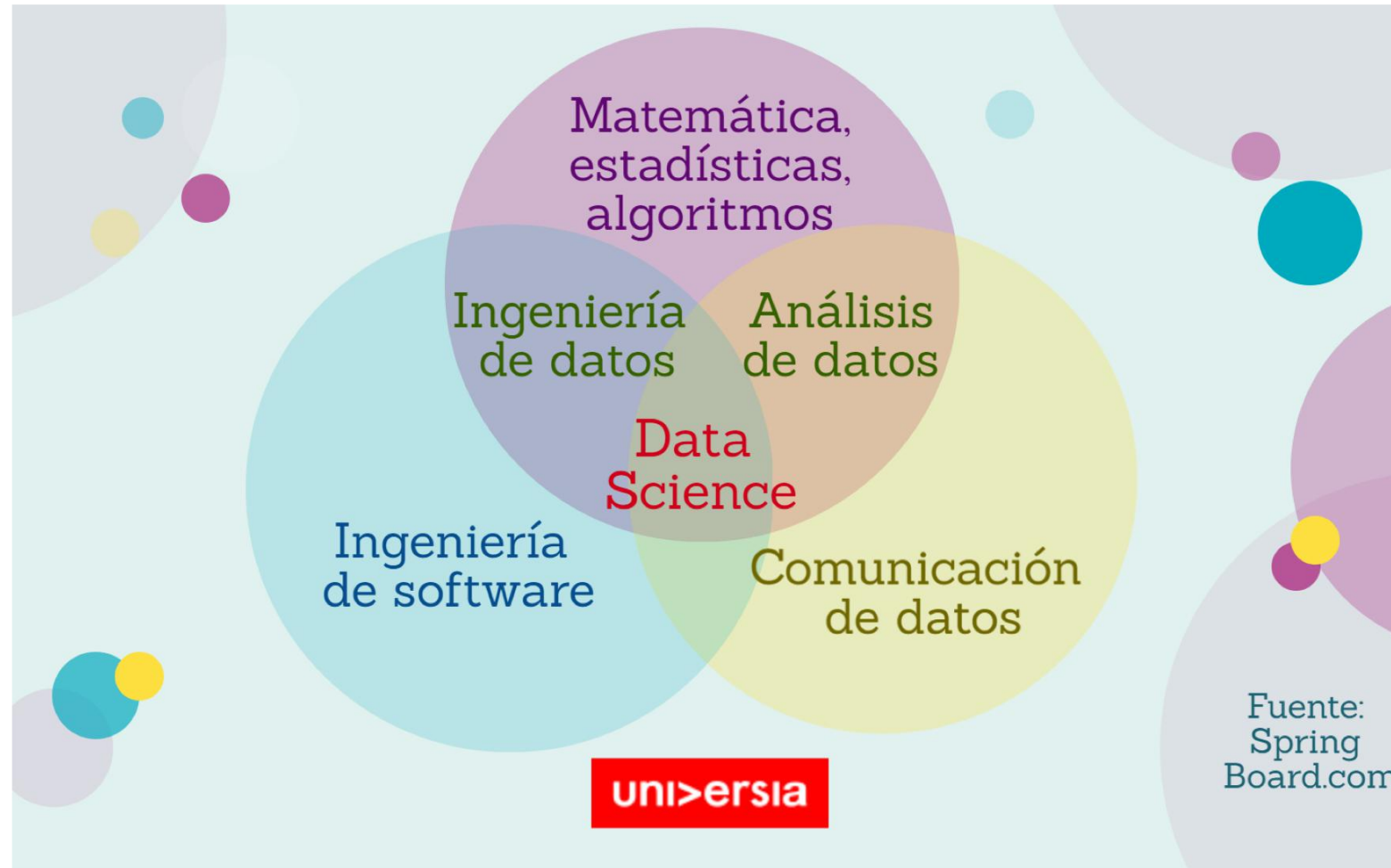




- Data Science es un **campo interdisciplinario** que involucra **métodos científicos, procesos y sistemas** para **extraer conocimiento** o un **mejor entendimiento de datos** en sus diferentes formas, ya sea **estructurados o no estructurados**, lo cual es una continuación de algunos campos de análisis de datos como la **estadística**, la **minería de datos**, el **aprendizaje automático** y la **analítica predictiva**.



# ¿Qué es Data Science?



#DataScienceEnPython

# ¿Qué es Data Science?

Es la **ciencia** que nos ayuda a responder preguntas a través del análisis de datos.



#DataScienceEnPython

# En nuestro grupo de *Meetup*

Buscamos **compartir conocimientos y experiencias**, hablar, **debatir** sobre Data Science y como Python y sus librerías nos **ayudan** en nuestros proyectos, investigaciones, ...



#DataScienceEnPython





# En nuestro grupo de *Meetup*

Os animo a **participar** activamente, con **preguntas, comentarios,** y por supuesto buscamos proyectos o investigaciones para **presentar en futuros meetups.**



#DataScienceEnPython

# Agenda

**19:00** Bienvenida y objetivos del grupo DataScience  
Python TheCUBE Madrid. **Carlos Picazo**

**19:15** Clasificación de cáncer de mama utilizando sklearn  
**Diego García Morate**

**19:40** Una aplicación práctica del data science: Prevención  
de incendios forestales. **Ernesto Padilla Martínez, David  
Márquez e Ignacio Rivera.**

**20:05** Mesa Redonda Q&A

**20:20** Networking: Refrescos y cervezas

**21:00** Fin de evento



**#DataScienceEnPython**



# Agenda

**19:00** Bienvenida y objetivos del grupo DataScience  
Python TheCUBE Madrid. **Carlos Picazo**

**19:15** Clasificación de cáncer de mama utilizando sklearn  
**Diego García Morate**

**19:40** Una aplicación práctica del data science: Prevención  
de incendios forestales. **Ernesto Padilla Martínez, David  
Márquez e Ignacio Rivera.**

**20:05** Mesa Redonda Q&A

**20:20** Networking: Refrescos y cervezas

**21:00** Fin de evento



#DataScienceEnPython







# Trabajo de fin de Módulo Data Science – MIOTI

## Prevención de Incendios Forestales

David Márquez | Ernesto Padilla | Ignacio Rivera

07/03/2018

# Objetivos

1. Estudiar la situación actual de los incendios en España y entender el contexto en el que realizamos este análisis de series temporales.
2. Estudiar el impacto económico y ambiental de los incendios en distintas Comunidades Autónomas de nuestro país.
3. Conocer la influencia de distintas variables en la generación de los incendios: Temperatura, tiempo de reacción, causas, especies.
4. Predecir el número de incendios del año siguiente a partir de la información de la serie temporal de años anteriores.
5. Crear un modelo que nos permita anticiparnos en la detección de incendios y optimizar los recursos con los que contamos a la hora de actuar.

## TRES PREGUNTAS A RESOLVER

¿Es cierto todo lo que se dice en los medios de comunicación sobre los incendios?

¿Podemos mejorar la prevención y actuación ante incendios a través de Data Science?

¿Cómo podría el Internet de las Cosas ayudar a luchar contra los incendios?

# Datos y Materiales



Los datos obtenidos proceden del Ministerio de Medio Ambiente de España e incluyen una serie histórica de incendios desde 1968 hasta 2015. Se incluyen dos bases de datos distintas:

1. Datos únicos de cada incendio y características generales: Fecha, Lugar, Extensión, Causa, Motivaciones, Tiempo de extinción, Medios de rescate, Temperatura, Gastos de gobierno, Muertos, Heridos, Cortes en electricidad y hogares...etc.
2. Datos particulares del impacto en montes concreto y características generales: Extensión de monte, superficie afectada, tipo de vegetación, árboles, pérdidas, localización. ...etc.

Para el análisis hemos considerado el último intervalo desde el 2000 al 2015 para hacer más sencillo el análisis y obtener mayor precisión.

\*Los datos se han complementado con un Dataset de Github realizado por la fundación Civio.

1. INTRODUCCIÓN Y CONTEXTO HISTÓRICO.
2. PREPROCESAMIENTO DE DATOS.
3. EVOLUCIÓN TEMPORAL DE INCENDIOS.
4. ANÁLISIS, VISUALIZACIÓN Y CORRELACIONES: TEMPERATURA, CAUSAS, PELIGROSIDAD, VEGETACIÓN, PÉRDIDAS, TIEMPO DE REACCIÓN, COMUNIDADES AUTÓNOMAS.
5. PREDICCIÓN DE SERIES TEMPORALES EN PROPHET.
6. MACHINE LEARNING & DEEP LEARNING.



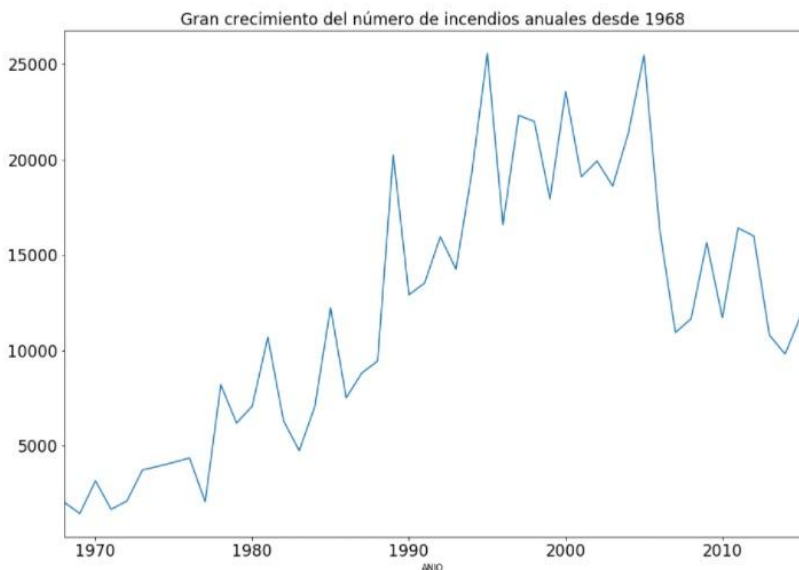
# Preprocesamiento

18 Chamaecyparis lawsoniana 21 Pinus sylvestris - Pinos 22 Pinus uncinata - Pinos 23 Pinus pinea - Pinos 24 Pinus halepensis - Pinos 25 Pinus nigra - Pinos 26 Pinus pinaster - Pinos 27 Pinus canariensis - Pinos 28 Pinus radiata - Pinos 31 Abies alba - Abeto 32 Abies pinsapo - Abeto 33 Picea excelsa - Pino 34 Pseudotsuga menziesii - Abeto de Douglas 35 Larix sp. 36 Cupressus sp. - Ciprés 37 Juniperus oxycedrus - Enebro 38 Junip.sabina phoenicea - Enebro 39 Otras coníferas 41 Quercus robur - Roble 42 Quercus petraea - Roble 43 Quercus pyrenaica - Roble 44 Quercus faginea - Roble 45 Quercus ilex - Roble/Alcornoque 46 Quercus suber - Alcornoque 47 Quercus rubra - Roble 49 Otros Quercus - Roble/Alcornoque 50 Populus x canadiensis - Álamos 51 Populus alba - Álamos 52 Populus - Álamo 53 Populus nigra - Álamo 54 Alnus glutinosa - Aliso común 55 Fraxinus sp. - Fresnos 56 Ulmus sp. - Olmos 57 Salix sp. 58 Otros Populus - Álamos 59 Otros Árboles de ribera 60 Ilex aquifolium - Acebo 61 Eucalyptus globulus - Eucalipto 62 Eucalyptus camaldulensis - Eucalipto 63 Erica arborea canaria - Eucalipto 64 Eucalyptus nittens - Eucalipto 65 Eucalyptus dalrympleana - Eucalipto 66 Olea europaea - Oliva 69 Otros Eucalyptus - Eucaliptos 71 Fagus sylvatica - Haya 72 Castanea sativa - Castaño 73 Betula sp. - Abedul 74 Corylus avellana - Avellano 75 Juglans regia - Nogal 76 Acer sp. - Arce 77 Tilia sp. - Tilos 78 Sorbus sp. - Serbal 79 Otras frondosas 81 Phoenix dactylifera - Palmera 82 Fayal - Brezal canario 83 Robinia pseudoacacia - Falsa Acacia 84 Tamarix Gallica - Taraje 99 Sin codificar en origen

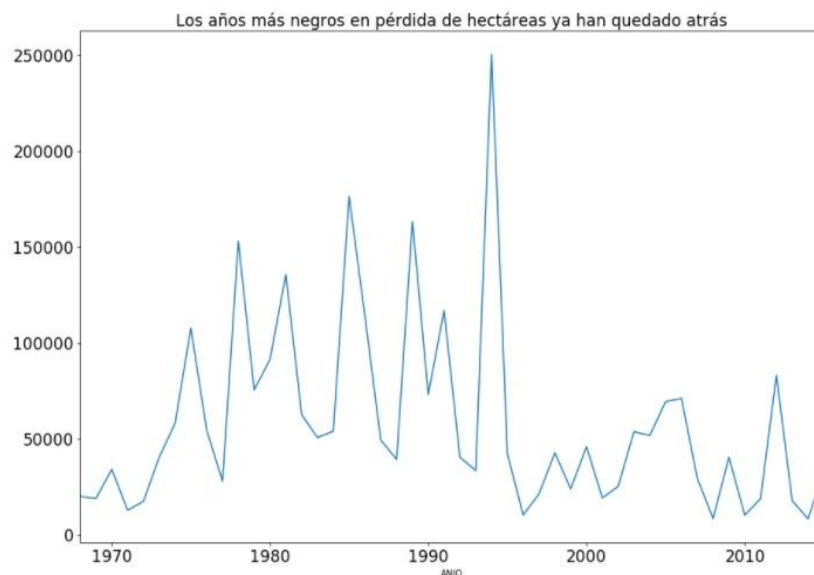
```
map_terrenos = {18:1, 35:1, 36: 1, 57:1, 59:1, 60:1,83:1,84:1, # Otros
                21 : 2, 22 : 2, 23: 2, 24:2, 25:2, 26:2, 27:2 ,28:2, 33: 2, # Pinos
                31:3 ,32:3, 34:3, # Abetos
                37:4, 38:4, # Enebro
                39:5, # Otras coníferas
                41:6, 42:6, 43:6, 44:6, 45:6, 46:6, 47:6, 48:6, 49:6, # Roble/Alcornoque
                50:7, 51:7,52:7,53:7,58:7, # Álamos
                54:8,# Aliso
                55:9, # Fresnos
                56:10, # Olmos
                61 : 11, 62:11,63:11, 64:11, 65: 11 ,69:11, # Eucalipto
                66: 12, # Oliva
                72: 13, # Castaño
                82: 14, # Brezal Canario
                99: 0
            }
map_terrenos_nombre = {1: 'Otros',
                       2:'Pinos',
                       3:'Abetos',
                       4:'Enebro',
                       5:'Otras coníferas',
                       6:'Roble/Alcornoque',
                       7:'Alamos',
                       8:'Alisos',
                       9:'Fresnos',
                       10:'Olmos',
                       11:'Eucaliptos',
                       12:'Olivas',
                       13:'Castaños',
                       14:'Brezal Canario'
                       }
```

- Importación CSVs
- Integración de bases de datos
- Conversión de tipología de datos
- Imputación de valores nulos y ausentes
- Eliminación de variables
- Gestión de diccionarios
- Agrupación de variables
- Preprocesamiento lógico y racional

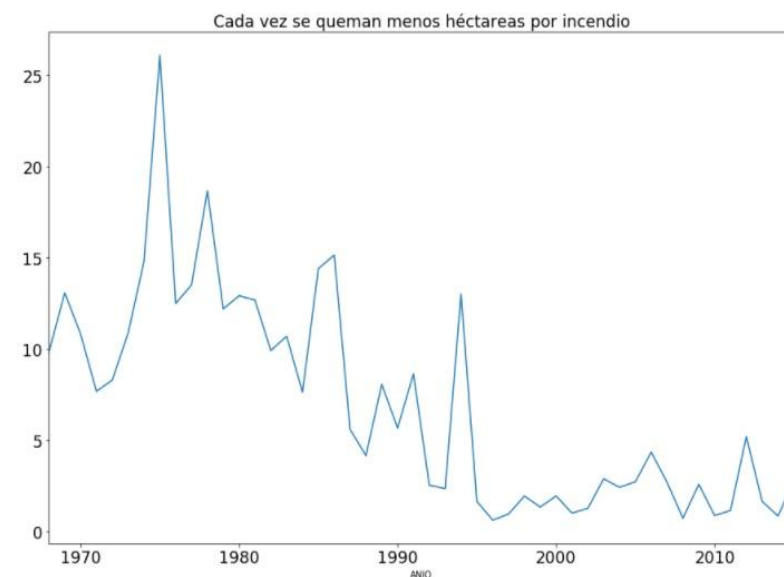
# Cada vez menos incendios graves



**EVOLUCIÓN DEL NÚMERO DE INCENDIOS  
POR AÑO**



**EVOLUCIÓN DEL NÚMERO DE HECTÁREAS  
QUEMADAS CADA AÑO**



**EVOLUCIÓN DEL NÚMERO DE HECTÁREAS  
QUEMADAS POR INCENDIO**

A PESAR DE QUE PUEDE PARECER QUE LA SITUACIÓN DE LOS INCENDIOS SE ESTÁ AGRAVANDO CON EL PASO DEL TIEMPO, LO QUE OBSERVAMOS ES UNA TENDENCIA DECRECIENTE EN LOS ÚLTIMOS AÑOS TANTO EN NÚMERO DE INCENDIOS COMO EN NÚMERO DE HECTÁREAS QUEMADAS POR INCENDIO.

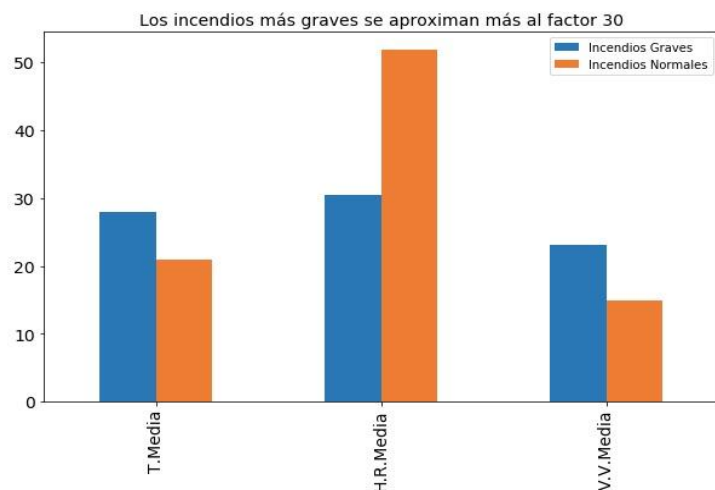
# Tres preguntas a responder

“El origen en el denominado Factor 30.”

SÓLO UN 0,13% DE TODOS LOS INCENDIOS SE HA GENERADO BAJO CONDICIONES DE FACTOR 30.

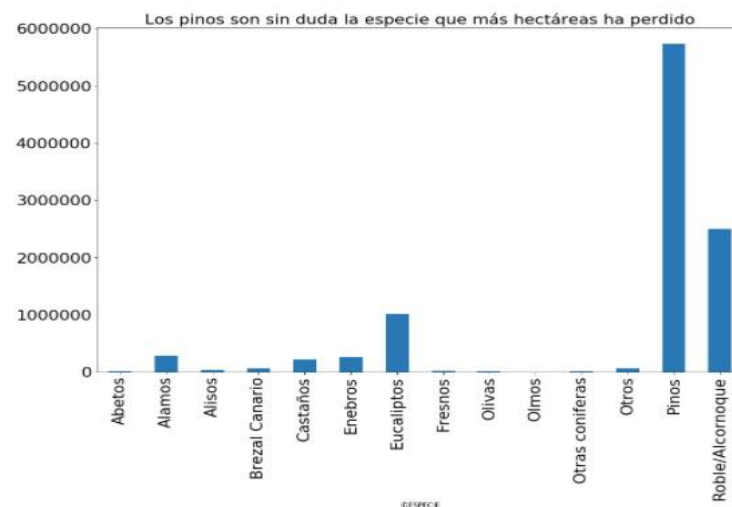
ESTE 0,13% DE LOS INCENDIOS REPRESENTA UN 1,54% DEL TOTAL HECTÁREAS QUEMADAS

LA MEDIA DE EXTENSIÓN DE INCENDIOS CON FACTOR 30 ES DE 101 HECTÁREAS POR 8 HECTÁREAS EN LA MEDIA TOTAL

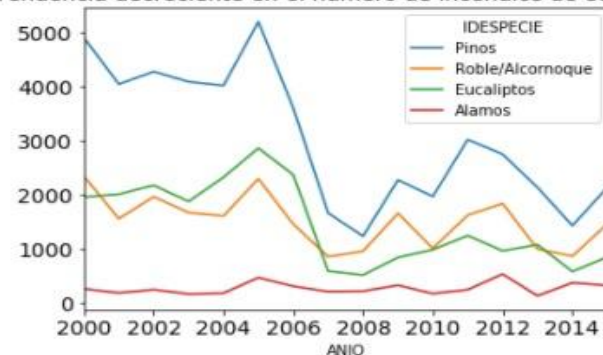


El factor 30 no genera incendios, los hace más devastadores.

“La culpa es de los Eucaliptos.”

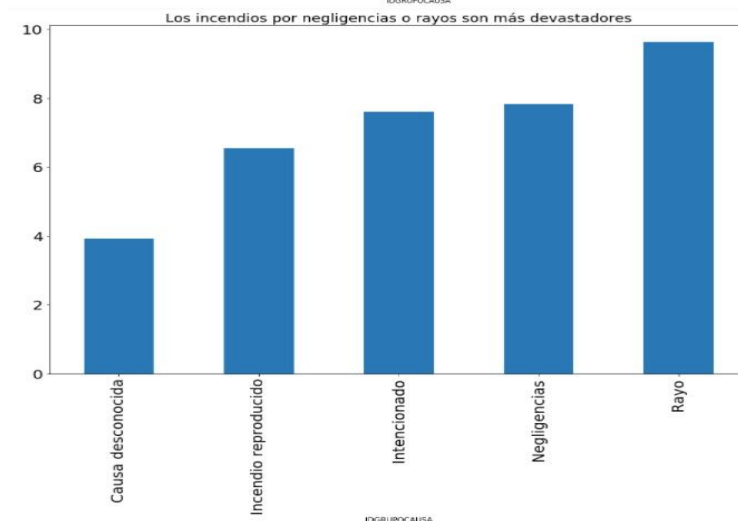
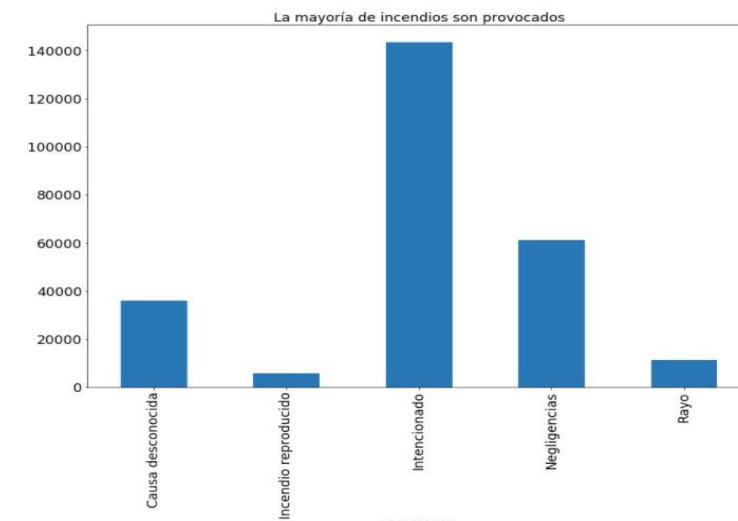


Tendencia decreciente en el número de incendios de eucaliptos



Hay árboles mucho más presentes en incendios que los eucaliptos.

“Los peores incendios son los provocados”



Los incendios por negligencias o rayos queman más hectáreas por incendio.

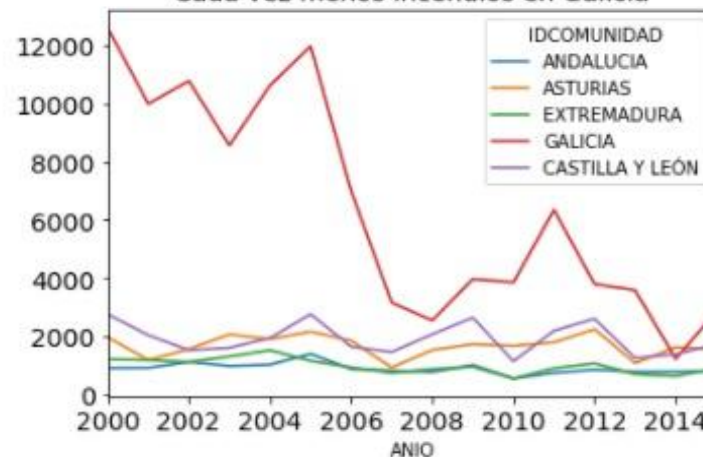


# Concentración en regiones concretas

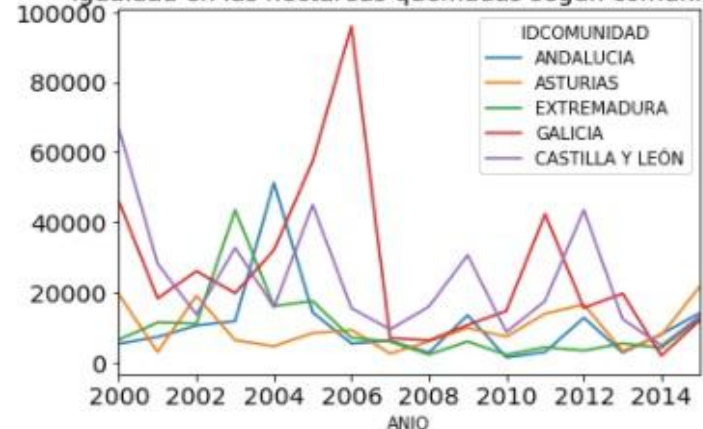
Incendios en España



Cada vez menos incendios en Galicia



Igualdad en las hectáreas quemadas según comunidad

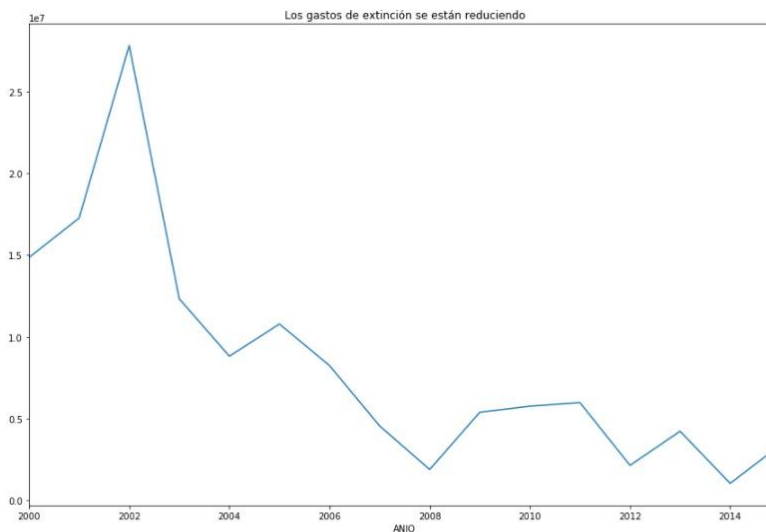
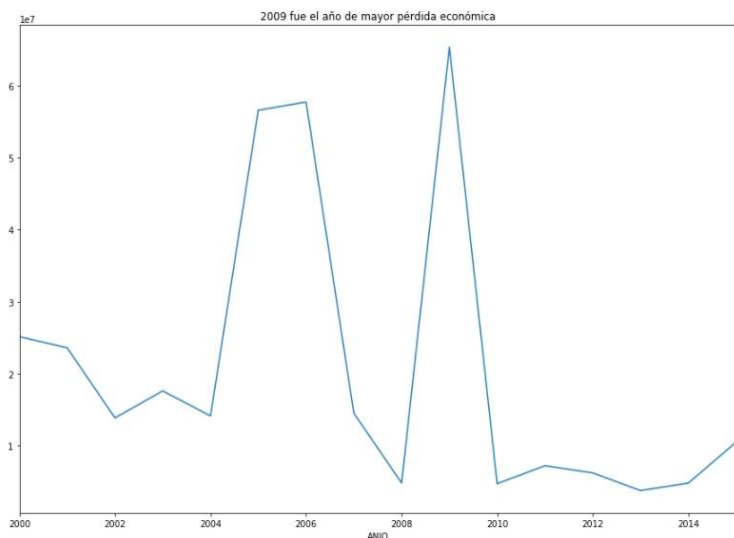


Hay 5 comunidades que destacan por encima del resto en cuanto a número de incendios y hectáreas quemadas. Galicia es sin duda la comunidad más perjudicada aunque en los últimos años la situación es más similar entre las distintas comunidades.

# Soluciones para una problemática real

Out[109]:

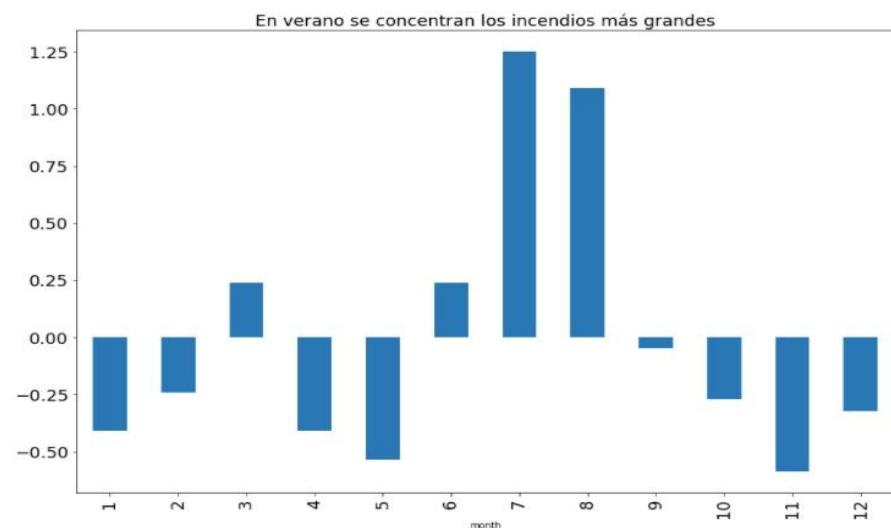
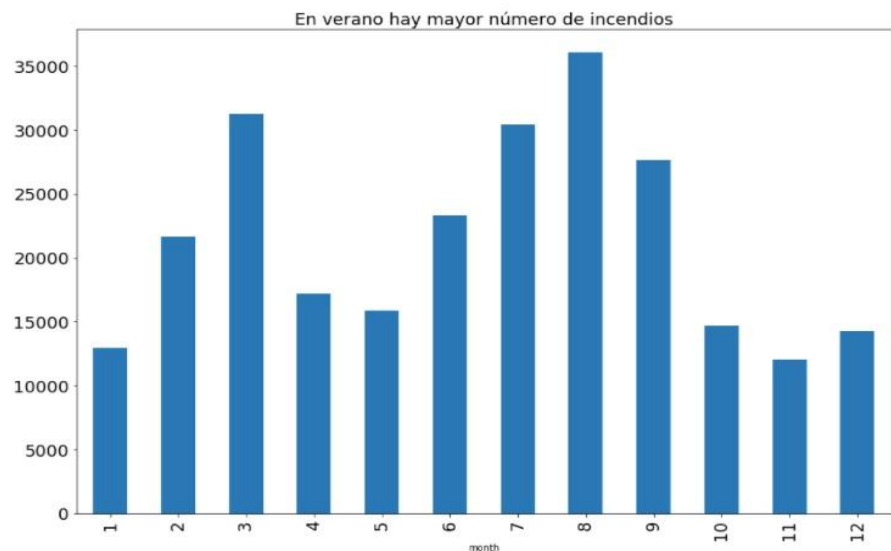
	IDPIF	DIFERENCIA_HORA_LLEGADA	SUPERFICIE_QUEMADA_TOTAL
224137	2012460248	00:25:00	28879.10
151720	2007350042	00:33:00	18672.90
100554	2004410050	00:16:00	18065.00
152869	2007380023	00:20:00	16820.63
109192	2005190182	00:33:00	12732.73
88520	2004210126	01:00:00	11802.00
214574	2012240531	01:08:00	11592.00
211690	2012120050	00:39:00	10613.25
105142	2005100838	00:11:00	9901.47
249388	2015230068	01:10:00	9756.76



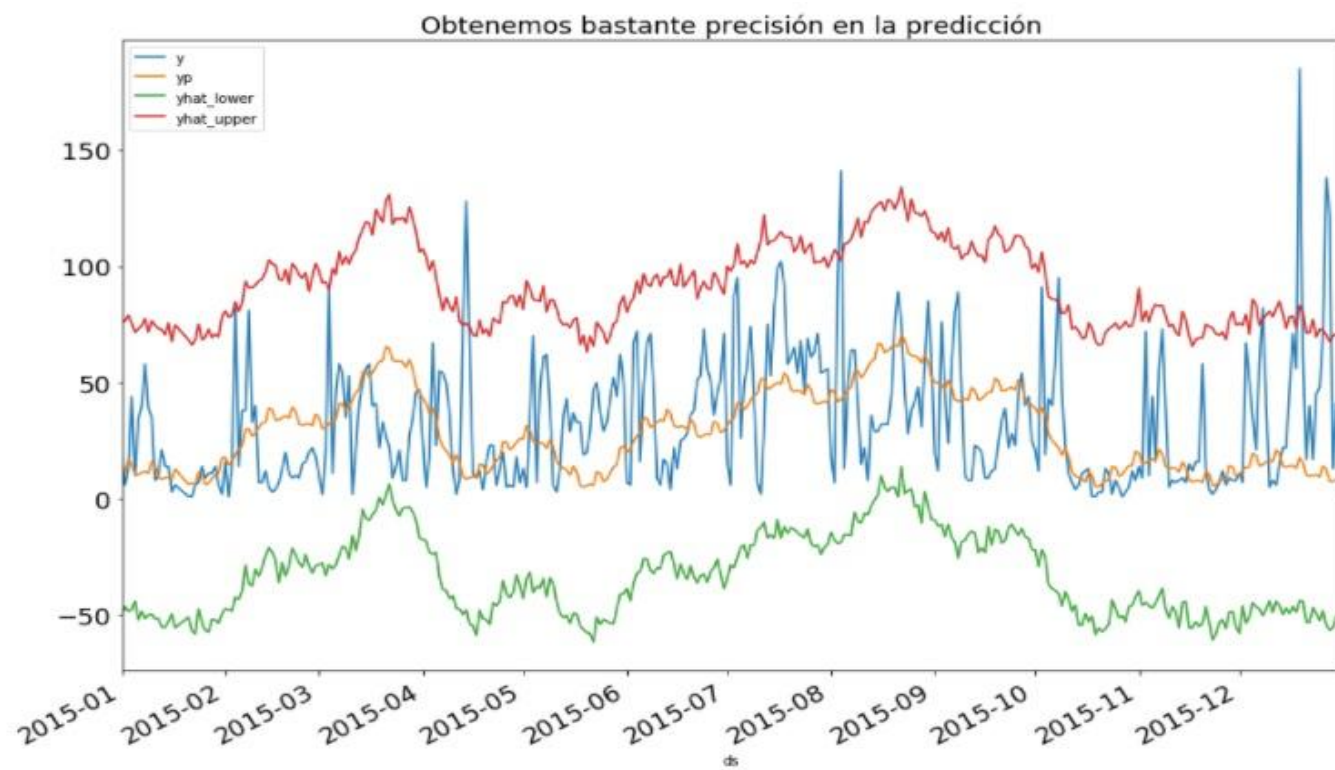
Tanto el estudio de los datos como las aplicabilidades y soluciones resuelven problemas reales a través de la tecnología:

- El tiempo de reacción ante los incendios tiene gran margen de mejora.
- Se solucionarían pérdidas económicas enormes para las comunidades autónomas.

# Predicción de series temporales



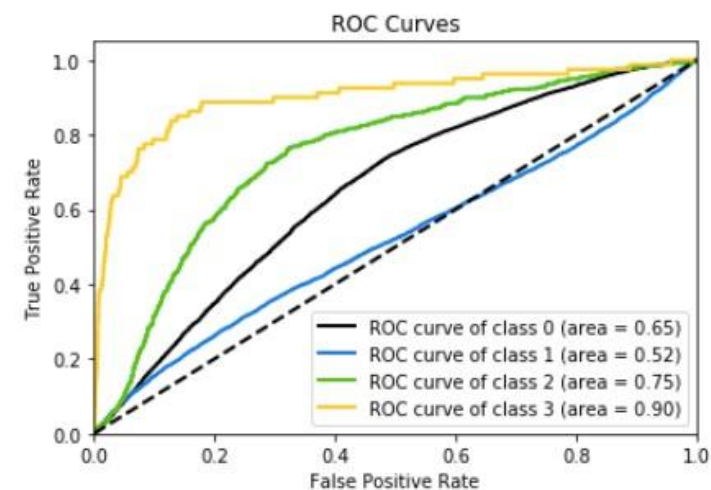
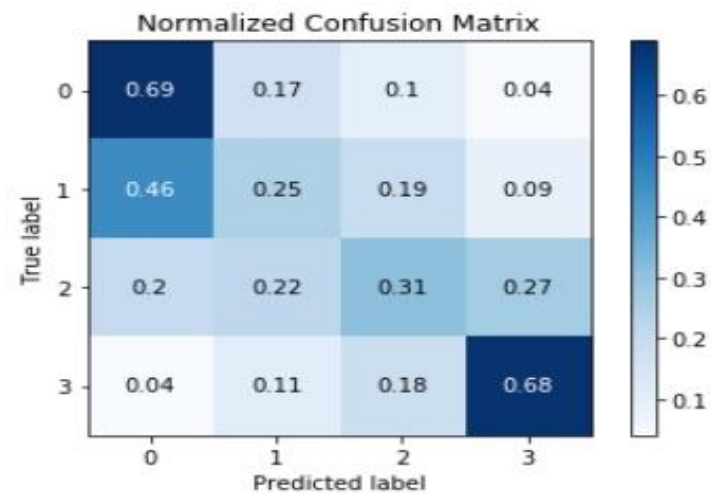
1. Nos permite conocer los focos de incendio antes de que se produzcan.
2. Optimización de recursos y dispositivos de extinción.





# Machine Learning

1. Limpieza del modelo y selección de las variables clave.
2. Definición de la variable objetivo de estudio / predicción.
3. Discretización de la variable superficie quemada por hectáreas.
  - 0: 0 hasta 1,5
  - 1: > 1,5 hasta 50
  - 2: > 50 hasta 500
  - 3: > 500
4. Flujo de creación de modelo predictivo:
  - Parámetro de evaluación.
  - Primera evaluación de clasificadores.
  - Análisis de resultados y problemas encontrados.
  - Resolución de problemas.
5. Modelo Final: Regresión Logística.
  - Newton – Cg.
  - Balanced weights.



# Solución tecnológica - IOT

#DataScienceEnPython



Después de todo el proyecto, hemos descubierto las claves detrás de los incendios en España y proponemos un modelo que podría ayudar a combatirlos.

- **PROBLEMA A RESOLVER:** Prevención, tiempo de reacción e información sobre el incendio.
- **DIRIGIDO A:** Gobiernos, reservas naturales, terrenos de alto valor, grandes complejos turísticos.
- **UTILIDADES:** Predicciones, diagnóstico de riesgo, selección de prioridades, simulación de expansión, actuación en tiempo real, optimización de recursos.
- **DESCRIPCIÓN:** Plataforma de información abierta y modelo de análisis de Machine Learning abierto a clientes e interesados para combatir los incendios en España.

+ MACHINE LEARNING



+ DATA SCIENCE

## Despliegue de red de sensores de meteorología y detección de gases.

Medición de temperatura, velocidad del viento, humedad y lluvia para calcular riesgo de incendio.



Detección de fuego a través de cambios en la composición del aire.



## Imágenes satelitales

Imágenes sobre la vegetación y su evolución en tiempo real.



## Plataforma de imágenes de Drones

Imágenes propias y compartidas: Estado del aire, cámara ultravioleta e infrarrojos calidad.



## Herramientas de predicción.

Optimización de recursos y aprovisionamiento previo a generación de incendios



# Conclusiones

1. Un contexto histórico marcado en la actualidad por una **menor presencia de incendios** al año y menor gravedad de los mismos.
2. Importancia de **explorar los datos** para entender el impacto de las distintas variables en la generación de incendios: Temperatura, tipo de vegetación, causalidad.
3. Concentración en **zonas concretas** de España donde se debería invertir en buscar posibles soluciones a través de la tecnología.
4. Hay distintas **variables que impactan en la propagación del incendio** y hemos tenido en cuenta para definir el modelo: Iniciación del incendio, tiempo de reacción, variables de temperatura, localización, tipo de vegetación y foco de origen, riesgo previo en la zona, estacionalidad.
5. Es posible realizar predicciones de series temporales para conocer el número de incendios esperados y podemos crear un **modelo que nos defina con suficiente precisión el riesgo** que puede tener un incendio generado.
6. Contando con todavía más variables e información, creemos que **el Data Science y el Internet de las cosas podrían ayudar** a combatir los incendios en España.





# Muchas gracias

*David Márquez: [davidmarquez.90@gmail.com](mailto:davidmarquez.90@gmail.com)*

*Ernesto Padilla: [epadillama89@gmail.com](mailto:epadillama89@gmail.com)*

*Ignacio Rivera : [Ignacioriveragg@gmail.com](mailto:Ignacioriveragg@gmail.com)*



 #DataScienceEnPython

*Meetup*

Q & A

 /THECUBEMadrid

 @THECUBEMadrid

 THECUBE Madrid

 /thecubemadrid

Sponsored by: 