

Taller de Exploración de datos Abiertos

Angela Devia

Líder de Desarrollo e Innovación, Científicas de Datos

www.angeladevia.com



Organiza:
UNIDAD ADMINISTRATIVA ESPECIAL DE REHABILITACIÓN Y MANTENIMIENTO VIAL - UMV
Secretaría Distrital de Movilidad

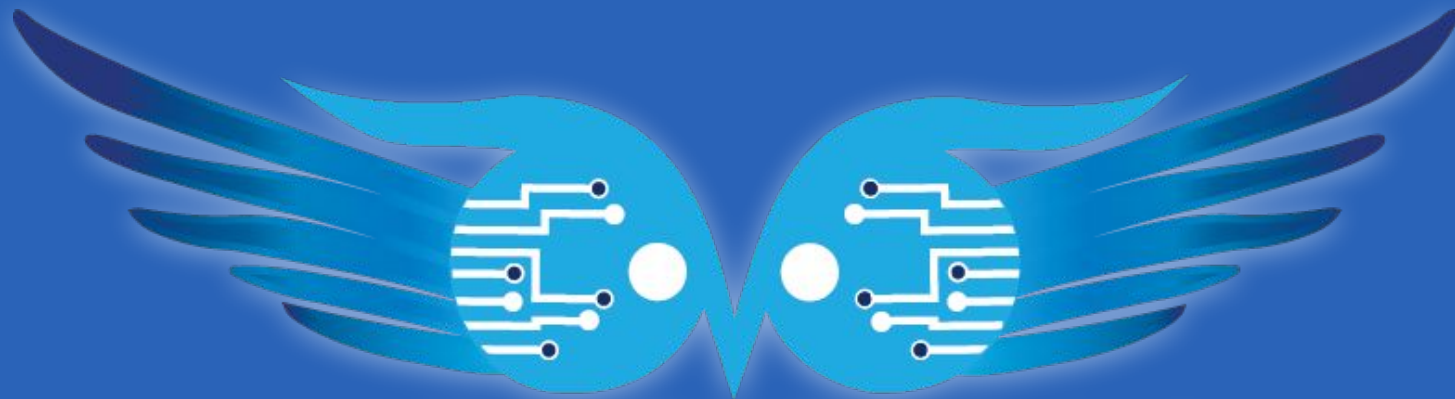
Agenda

Introducción a la Ciencia de datos

¿Cómo se usa para la toma de decisiones?

Exploración de datos





CIENTIFICAS DE DATOS

contacto@cientificasdedatos.com

<https://www.facebook.com/cientificasdedatos/>

Introducción a la Ciencia de Datos

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

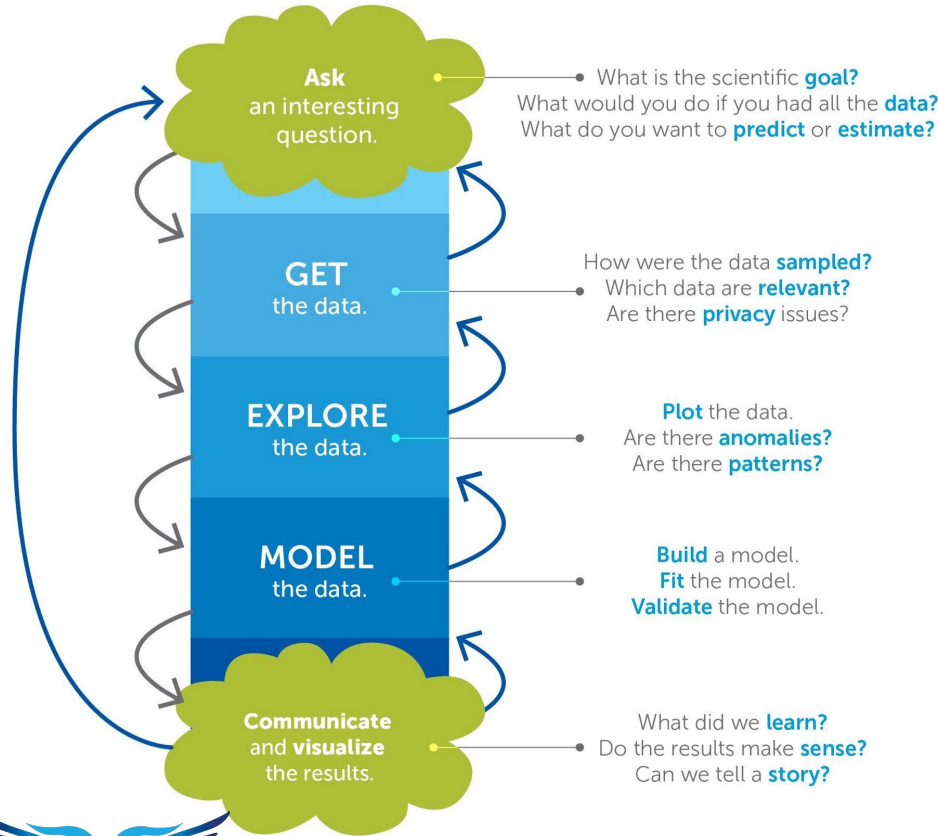
- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



The Data Science Process



Hagamos Ciencia de Datos

Datos:
<https://goo.gl/9WTUrV>

Mapa:
<https://goo.gl/EfevLd>

Resumen:
<https://goo.gl/kqVWDW>

¿Cómo se usa para la toma de decisiones?



Haciendo una pregunta interesante...

Hipótesis formuladas

Relación de longitud, ancho y área por localidad, ¿influye en el tiempo de respuesta?

¿Influye la ubicación en el tiempo de reparación de la malla vial?

¿Existe “Discriminación por localidad”?

¿El equipo, personal y contratista seleccionado tienen incidencia en el mantenimiento de la malla vial?

¿A dónde queremos llegar?

Diseñar modelos predictivos para mejorar la productividad de la entidad

Mantenimiento por petición

Mantenimiento predictivo

Técnicas:

APRENDIZAJE

SUPERVISADO

- Clasificación
- Regresión

APRENDIZAJE NO

SUPERVISADO

- Agrupamiento
- Asociación
- Sumarización

¿Qué podemos hacer para construir un modelo?

Podemos utilizar técnicas de aprendizaje de máquina

Hacer análisis y formular preguntas a partir de nuestro conocimiento

Soluciones alrededor del mundo

Inteligencia artificial en accidentes



Inteligencia Artificial: RoadDoctor



Inteligencia Artificial: Tráfico en China



Caso: Uber Eats



Herramientas para información geográfica



...

Exploración de datos espaciales



Funcionalidades a probar:

- Carga de información geográfica
- Análisis de la información

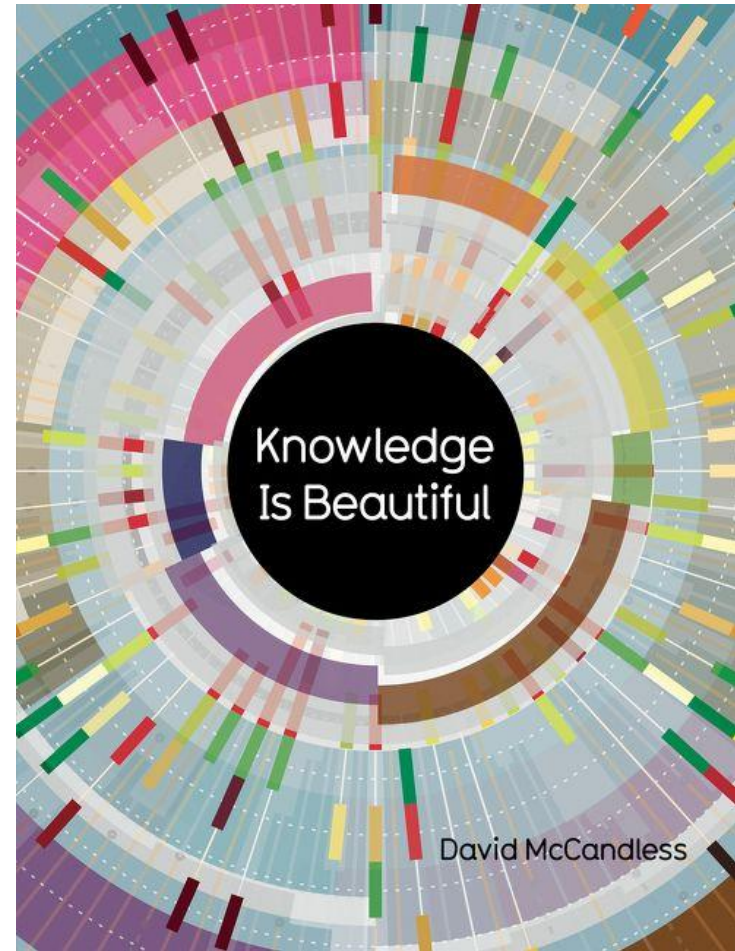
Visualización de datos

Librerías de R como:

[Geoplumber](#)

[Geopandas](#)

Javascript. Ejemplo: [BioTablero](#)



Conclusiones del taller



Gracias por su
atención y
participación

Cualquier inquietud, contáctenos:
contacto@cientificasdedatos.com



Scan me



2C9U W6

