



# Ingeniería de Características

Semestre 2022-2

Julio Weissman

[julio.waissman@unison.mx](mailto:julio.waissman@unison.mx)

# Ya no estamos a distancia

---



Clases  
presenciales



Mucho trabajo  
autónomo



Uso de muchas  
herramientas

# Que herramientas vamos a usar

[Teams](#) para los cursos presenciales a distancia, archivos, encuestas y evidencias

[Página](#) de Github del curso con material y enlaces

[DataCamp](#), [EdX](#), [Coursera](#) y otros MOOCs para partes de contenido de cursos

[Kaggle](#) para datos, problemas y actividades de aprendizaje

[Slack](#) para mantener la comunicación

[Github](#) para los trabajos prácticos

[Medium](#) para lectura de documentos y escritura de blogs

# Como vamos a trabajar

---

## **martes**

Dos horas el martes (6 a 8 p.m.) con el tema nuevo de la semana

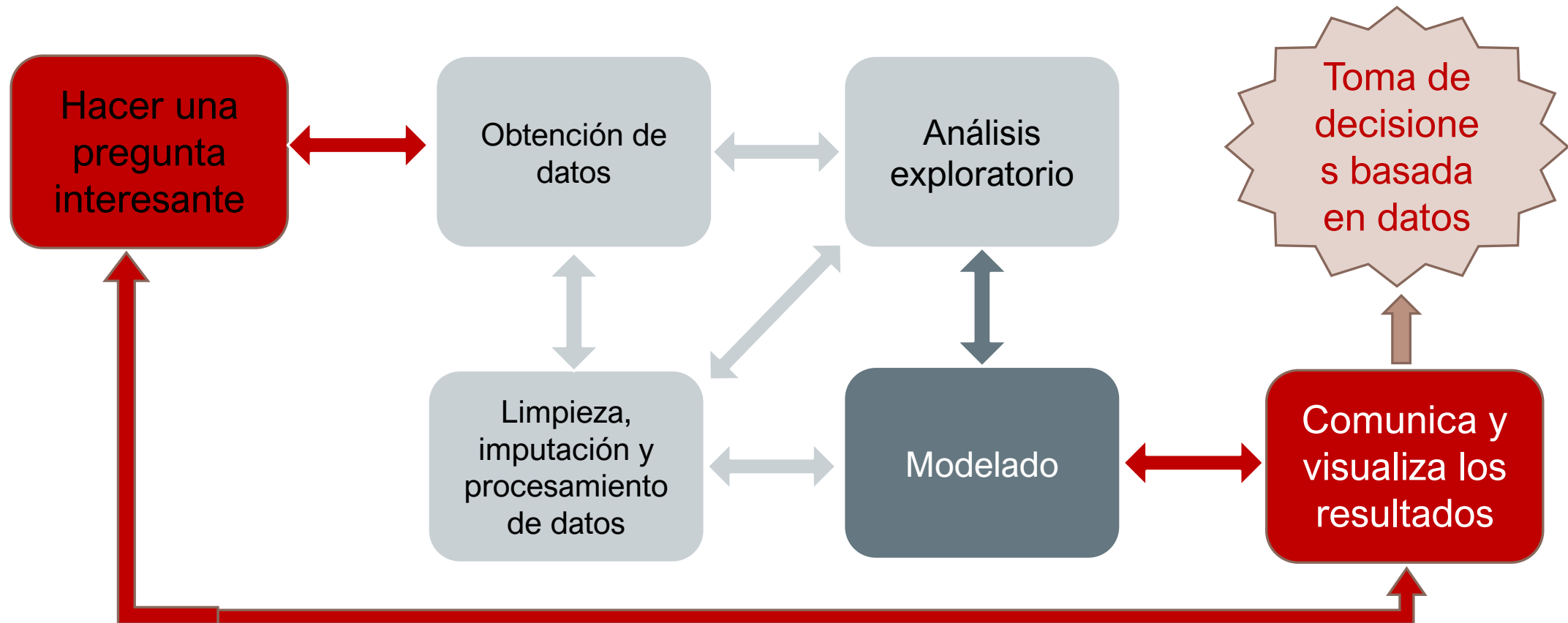
## **sábado**

Dos horas el sábado (10 a 12 p.m.) para aspectos prácticos, cerrar el tema y tópicos complementarios

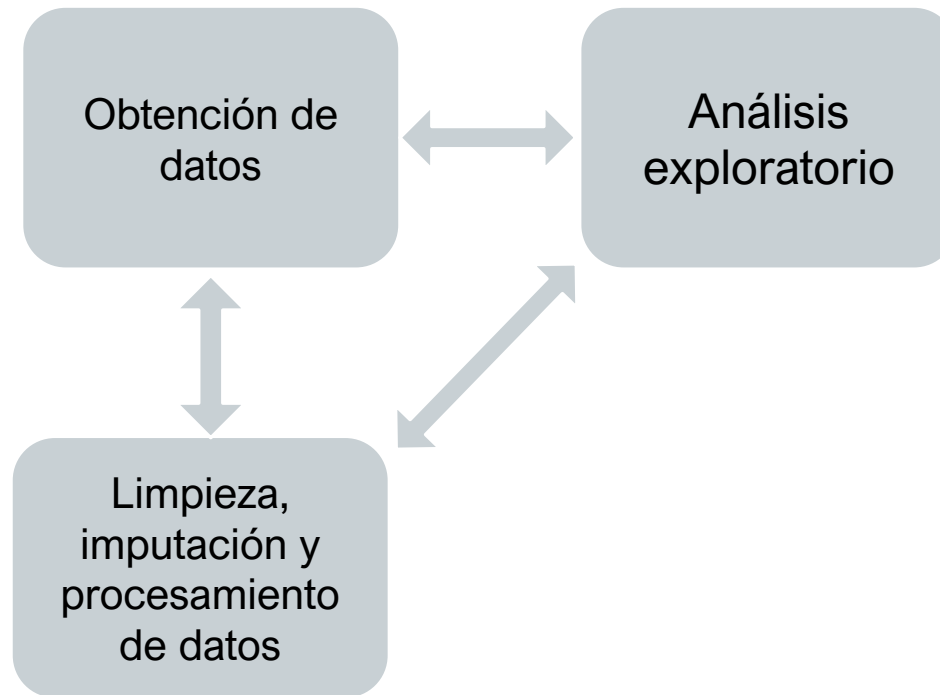
## **jueves y viernes**

Tres horas de trabajo independiente para avanzar en las actividades asignadas el martes (asesorías incluidas)

# Motivación



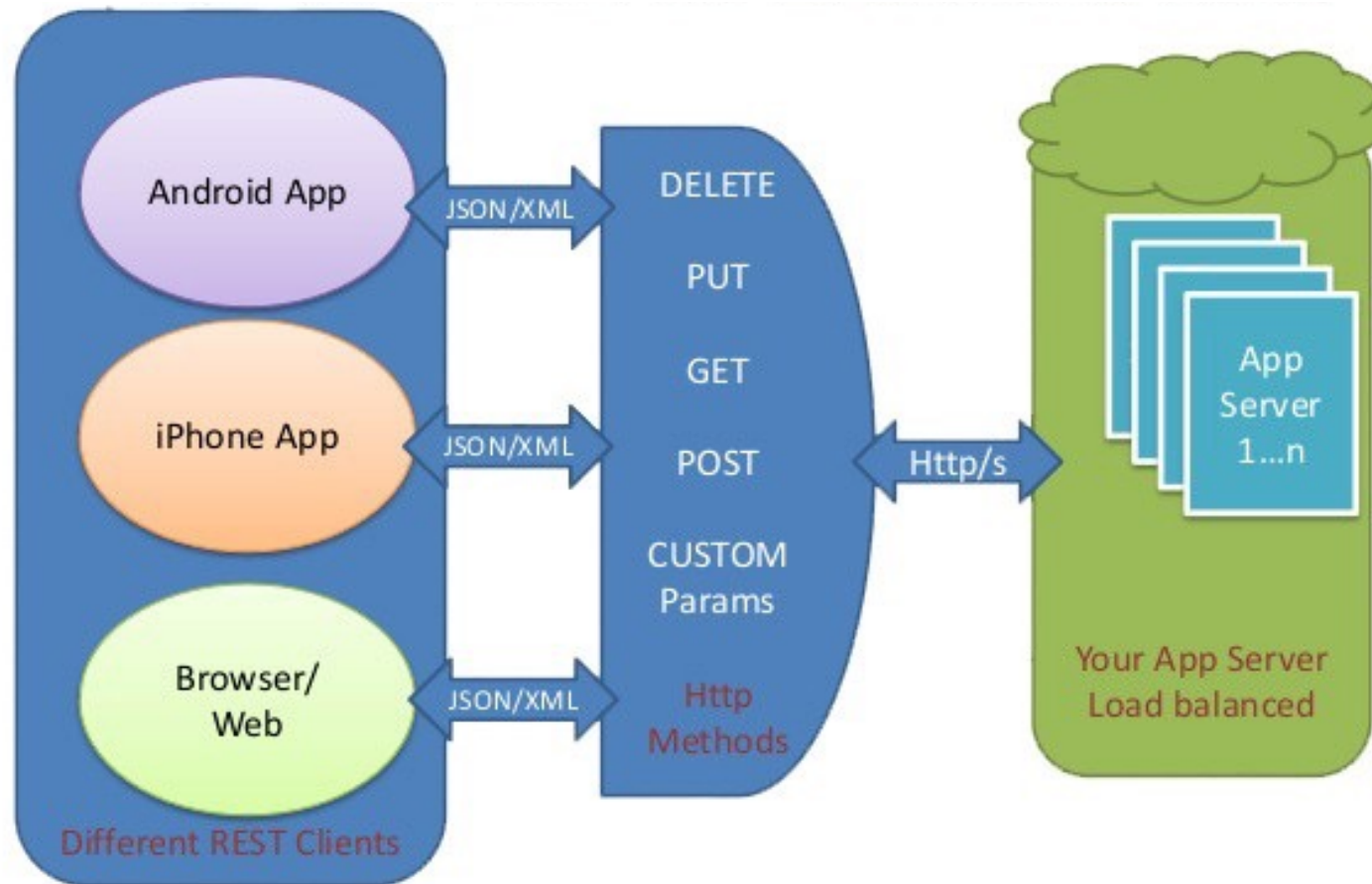
# Motivación



# Como queremos los datos

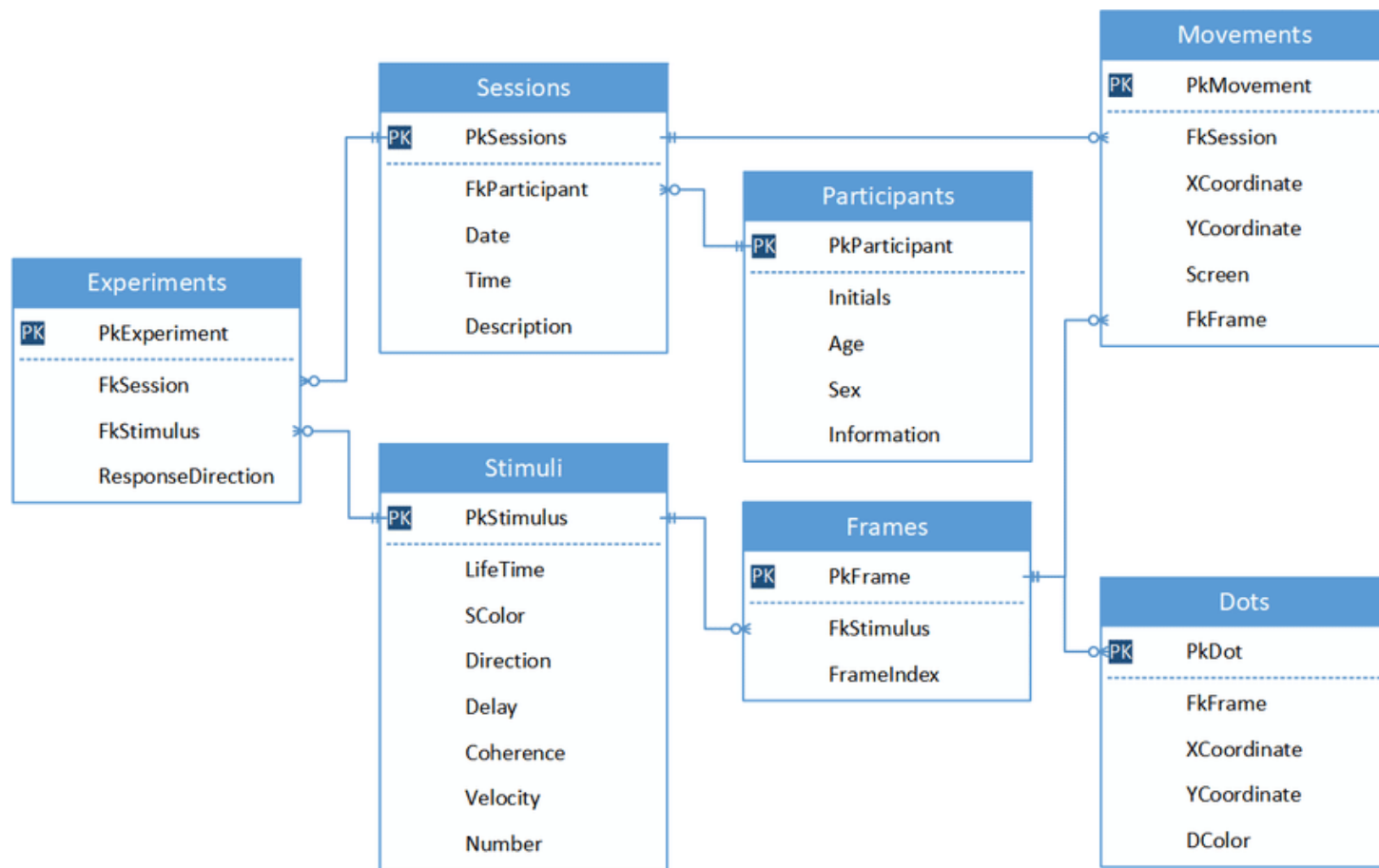
	UID	iso2	iso3	code3	FIPS	Admin2	Province_State	Country_Region	Lat	Long_	Combined_Key	1/22/20	1/23/20	1/24/20	1/25/20
1	16	AS	ASM	16	60		American Samoa	US	-14.27100	-170.13200	American Samoa, US	0	0	0	0
2	316	GU	GUM	316	66		Guam	US	13.44430	144.79370	Guam, US	0	0	0	0
3	580	MP	MNP	580	69		Northern Mariana Islands	US	15.09790	145.67390	Northern Mariana Islands, US	0	0	0	0
4	63072001	PR	PRI	630	72001	Adjuntas	Puerto Rico	US	18.18012	-66.75437	Adjuntas, Puerto Rico, US	0	0	0	0
5	63072003	PR	PRI	630	72003	Aguada	Puerto Rico	US	18.36025	-67.17513	Aguada, Puerto Rico, US	0	0	0	0
6	63072005	PR	PRI	630	72005	Aguadilla	Puerto Rico	US	18.45968	-67.12082	Aguadilla, Puerto Rico, US	0	0	0	0
7	63072007	PR	PRI	630	72007	Aguas Buenas	Puerto Rico	US	18.25162	-66.12681	Aguas Buenas, Puerto Rico, US	0	0	0	0
8	63072009	PR	PRI	630	72009	Aibonito	Puerto Rico	US	18.13136	-66.26413	Aibonito, Puerto Rico, US	0	0	0	0
9	63072011	PR	PRI	630	72011	Anasco	Puerto Rico	US	18.28798	-67.12061	Anasco, Puerto Rico, US	0	0	0	0
10	63072013	PR	PRI	630	72013	Arecibo	Puerto Rico	US	18.40663	-66.67508	Arecibo, Puerto Rico, US	0	0	0	0
11	63072015	PR	PRI	630	72015	Arroyo	Puerto Rico	US	17.99846	-66.05655	Arroyo, Puerto Rico, US	0	0	0	0
12	63072017	PR	PRI	630	72017	Barceloneta	Puerto Rico	US	18.44553	-66.56053	Barceloneta, Puerto Rico, US	0	0	0	0
13	63072019	PR	PRI	630	72019	Barranquitas	Puerto Rico	US	18.20159	-66.30963	Barranquitas, Puerto Rico, US	0	0	0	0
14	63072021	PR	PRI	630	72021	Bayamon	Puerto Rico	US	18.34946	-66.16844	Bayamon, Puerto Rico, US	0	0	0	0
15	63072023	PR	PRI	630	72023	Cabo Rojo	Puerto Rico	US	18.04099	-67.15439	Cabo Rojo, Puerto Rico, US	0	0	0	0
16	63072025	PR	PRI	630	72025	Caguas	Puerto Rico	US	18.21161	-66.05078	Caguas, Puerto Rico, US	0	0	0	0
17	63072027	PR	PRI	630	72027	Camuy	Puerto Rico	US	18.41858	-66.86021	Camuy, Puerto Rico, US	0	0	0	0
18	63072029	PR	PRI	630	72029	Canovanas	Puerto Rico	US	18.32880	-65.88761	Canovanas, Puerto Rico, US	0	0	0	0
19	63072031	PR	PRI	630	72031	Carolina	Puerto Rico	US	18.37499	-65.95683	Carolina, Puerto Rico, US	0	0	0	0
20	63072033	PR	PRI	630	72033	Catano	Puerto Rico	US	18.43727	-66.14331	Catano, Puerto Rico, US	0	0	0	0
21	63072035	PR	PRI	630	72035	Cayey	Puerto Rico	US	18.10285	-66.14914	Cayey, Puerto Rico, US	0	0	0	0
22	63072037	PR	PRI	630	72037	Ceiba	Puerto Rico	US	18.25182	-65.66642	Ceiba, Puerto Rico, US	0	0	0	0
23	63072039	PR	PRI	630	72039	Ciales	Puerto Rico	US	18.28885	-66.51647	Ciales, Puerto Rico, US	0	0	0	0
24	63072041	PR	PRI	630	72041	Cidra	Puerto Rico	US	18.17356	-66.16118	Cidra, Puerto Rico, US	0	0	0	0

# Como vienen los datos





# Como vienen los datos



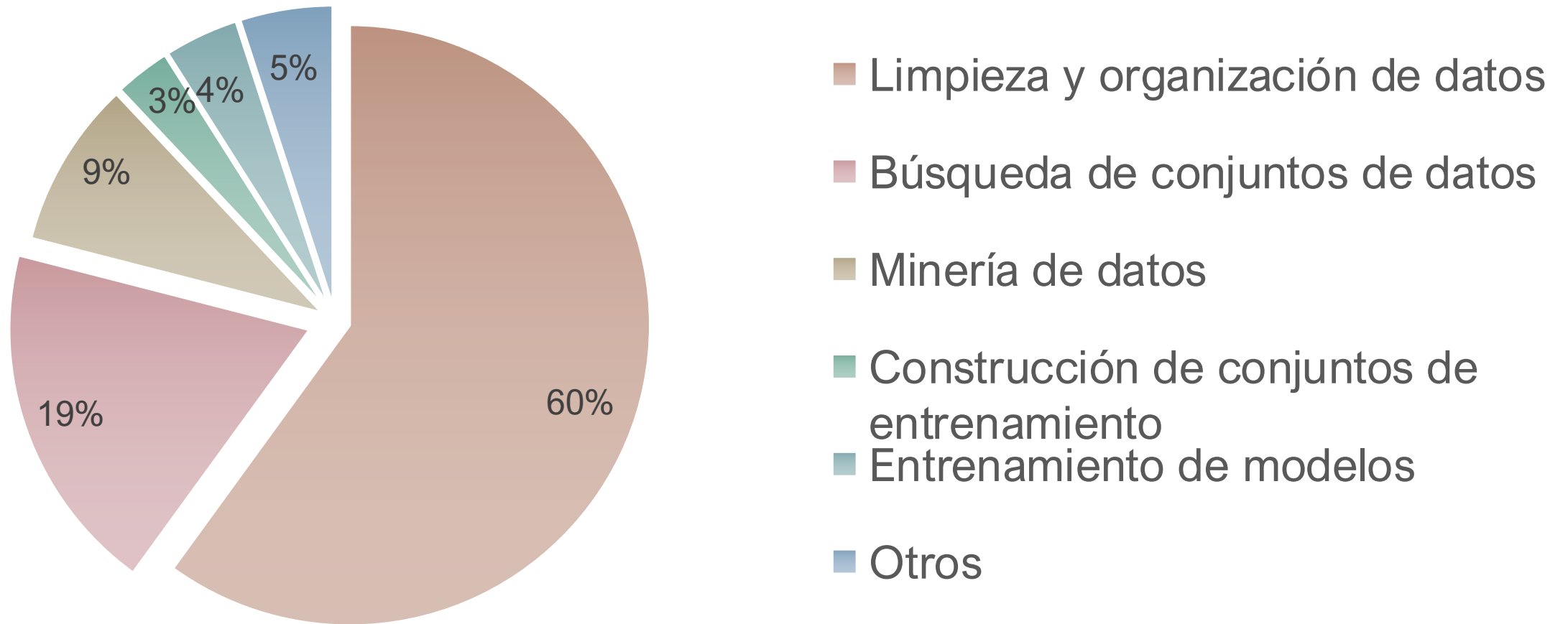
# Como vienen los datos

```
<page>
  <title>Julia Morilla de Campbell</title>
  <ns>0</ns>
  <id>4949229</id>
  <revision>
    <id>102979192</id>
    <parentid>100596558</parentid>
    <timestamp>2017-10-30T07:46:11Z</timestamp>
    <contributor>
      <username>InternetArchiveBot</username>
      <id>4704851</id>
    </contributor>
    <comment>Rescatando referencia 1 y marcando 0 como roto #IABot (v1.6)</comment>
    <model>wikitext</model>
    <format>text/x-wiki</format>
    <text xml:space="preserve" bytes="4165">'''Julia Morilla de Campbell''' ([[Rosario (Argentina)|Rosario]]
Nació en [[Rosario (provincia de Santa Fe)|Rosario]] ([[provincia de Santa Fe]]), hija de José María Morilla
Desde pequeña estudió canto varios años. En su adolescencia cantaba [[sevillana (música)|sevillanas]] y [[copl
```

## 11

F21													
	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	English												
4	Area:												
5	Customer's Task: sec												
6	OEE:												
7	Target Cycle Time: 0.0 sec												
8	Tolerance Time: 0.0 sec												
9													
10	Current State												
11	after Waste Elimination												
12	Operator	Station	Number	Operation	Time (s)	Code	Total time per Op.	Potential (s)	Potential per Op.	Remark	Time (s)	Total time per Op.	Operator
13	OP1			LOAD	5.0	va	57.3		6.0		5.0	#VALUE!	OP1
14	OP1			WIND	20.0	wa					20.0		OP1
15	OP1			INSTALL	8.0	va					8.0		OP1
16	OP1			WIND	25.0	va					25.0		OP1
17	OP1			TE	8.0	va					8.0		OP1
18	OP1			UNPACK	7.0	wa					7.0		OP1
19	OP1			MOVE	1.0	va					1.0		OP1
20	OP1			end							#VALUE!		OP1
21	OP1										#VALUE!		OP1
22	OP1										#VALUE!		OP1
23	OP1										#VALUE!		OP1
24	OP1										#VALUE!		OP1
25	OP1										#VALUE!		OP1
26	OP1										0.0		OP1
27	OP1										0.0		OP1
FROM SUBTITLE Changeover Analysis OUTPUT RAW DATA Backlog Current State after Waste Elimination													
Ready													

# ¿Qué hace un científico de datos?





## Objetivo del curso

# Temario

## Integracion de datos de diferentes fuentes

- Lectura de formatos csv, json, xml, sql.
- Uso de APIs

## Limpieza de datos

- Datos numéricos
- Datos cualitativos
- Cadenas de caracteres
- Series de tiempo

# Temario

## Transformación de datos

- Valores perdidos e imputación de datos
- Agregación
- Generación de características

## Análisis exploratorio de datos

- Indicadores claves de desempeño
- Herramientas para EDA
- Principios básicos de visualización
- Desarrollo de tableros para EDA

# Temario

## Casos especiales

- Información georeferenciada
- Series de tiempo
- Detección de anomalías

## Reducción de datos

- Selección de características
- Análisis en componentes principales
- Otras técnicas de reducción



# Criterios de evaluación



Participación (10%)



Evidencias de aprendizaje (40%)



Proyectos (50%)