



# DATOS 360°

Una visión integral para la gestión  
de los datos en el Estado



## Jornadas Tecnológicas

# Componente para análisis exploratorio de datos

Fernando Telechea  
arquitectura@agesic.gub.uy

# Agenda


- **¿A qué nos referimos con Análisis Exploratorio de Datos (AED)?**
- **Concepto de Cuaderno**
- **Herramientas para realizar AED**
- **Demo**

# Análisis Exploratorio de Datos

- El Análisis Exploratorio de Datos es un enfoque para el análisis de datos que emplea una variedad de técnicas (por lo general gráficas) para:
  - Maximizar el conocimiento sobre un conjunto de datos (dataset o dataframe)
  - Detectar valores atípicos y anomalías
  - Extraer variables importantes
  - Probar supuestos

# Cuadernos

- **U**
- **in**
- **co**
- **T**
- **re**
- **e**





localhost:8888/notebooks/mySample.ipynb

Más visitados Comenzar a usar Fir... De Google Chrome SP XX 2019 Capacida...

jupyter mySample Last Checkpoint: hace 5 minutos (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

 **agesic**  
DESARROLLANDO  
EL URUGUAY DIGITAL

 **PRESIDENCIA**  
REPÚBLICA ORIENTAL DEL URUGUAY

**Cuaderno de Ejemplo**

Esto es una celda explicativa  
de tipo Markdown

1	Stanford	96	151	55
2	Harvard	112	165	53
3	U.Penn	92	141	49
4	Princeton	90	137	47

In [ ]:

ite  
rkdown) y  
ación.

nuestran

# Herramientas - Proyecto Jupyter



jupy

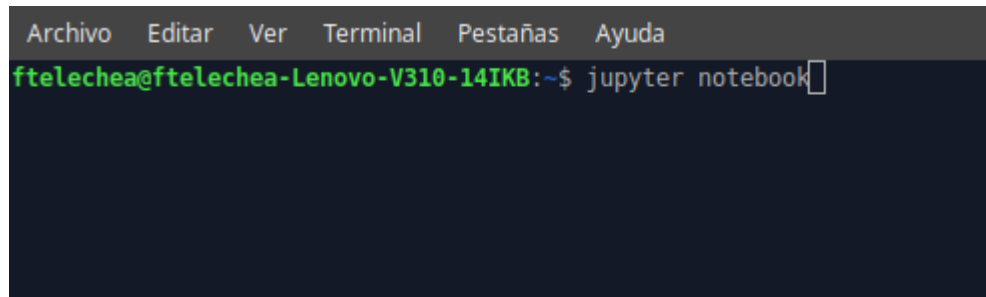
Jupy  
herr  
trab  
inte  
está

vidor  
(Tornado)  
que  
iones de  
os

# Crear Cuadernos

Dependiendo de la herramienta a utilizar se accede de distinta forma:

- Cuadernos locales: comando “jupyter notebook” o “jupyter lab”



```
Archivo  Editar  Ver  Terminal  Pestañas  Ayuda
ftelechea@ftelechea-Lenovo-V310-14IKB:~$ jupyter notebook
```

- Cuadernos en servidor: Se accede a la url

# Compartir Cuadernos



JUPYTER

FAQ



```
#pd.pivot_table(df_filter,index=["Descripcion_Inciso"], values=["monto_adj"], aggfunc='sum')

df_PorIncisos = df_view_final.groupby(['Descripcion_Inciso'])['monto_usd'].agg('sum')
df_PorIncisos = df_PorIncisos.reset_index()

#ordenar dataframe
df_PorIncisosSort = df_PorIncisos.sort_values('monto_usd', ascending=False)
df_PorIncisosSort
```

Out[9]:

	Descripcion_Inciso	monto_usd
8	Administración Nacional de Educación Pública	10367724742.25
4	Banco de Seguros del Estado	565879292.36
28	Instituto del Niño y Adolescente del uruguay	301032371.52
2	Banco Central del Uruguay	52786248.02
41	Ministerio del Interior	44170170.60
9	Administración de los Servicios de Salud del E...	43780061.55
36	Ministerio de Salud Pública	36695704.53
30	Ministerio de Defensa	31947940.09
5	Banco de la República Oriental del Uruguay	21405670.76
31	Ministerio de Economía y Finanzas	15476208.89

```
from plotly.offline import init_notebook mode
```

# Herramientas de Cuadernos “On-Line”



kaggle



Azure





# Referencias

- Para compartir notebooks a partir de github
  - <https://nbviewer.jupyter.org/>
- Ejecutar cuadernos Online
  - <https://mybinder.org/>
- Plataformas Online de Cuadernos
  - <https://jupyterhub.agesic.gub.uy> (ambiente Piloto, va a cambiar la url)
  - <https://colab.research.google.com> (uso libre CPU / GPU)
  - <https://www.kaggle.com/kernels> (uso libre)
  - <https://notebooks.azure.com/#jupyter>
- Principales bibliotecas utilizadas
  - Pandas - Extensión de NumPy para manipulación y análisis de datos (<https://pandas.pydata.org/>)
  - Plotly - Herramienta de visualización basada en D3.js (<https://plot.ly>)

# Alcance de la Demo

- Herramienta a utilizar: Jupyter Hub
- Lenguaje: Python
- Descripción rápida del dataframe
  - número de filas, columnas
  - tipos de datos
  - vista previa
- Distribución y correlación de datos
- Agregar campos calculados al dataframe
- Análisis de datos
  - group by
  - pivot table
  - gráficos