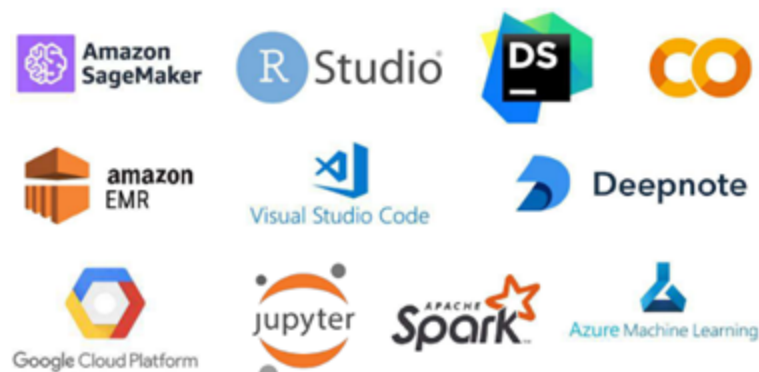


# Herramientas de software para el análisis exploratorio de datos

Es importante para tener un entorno en donde ejecutar los proyectos y donde te sientas más cómodo. Algunos ejemplos son:

- Amazon Sage Maker
- RStudio
- Google Cloud Platform
- Jupyter
- Visual Studio Code
- Deepnote
- Azure Machine Learning
- Apache Spark
- Amazon EMR



## Jupyter notebooks

Son bastante populares debido a que te permiten combinar texto con código.

Hay celdas de código llamadas **chunks**, que te ayudan a ir construyendo tu código y así hacer un prototipo de lo que quieres lograr.

## RStudio

Te permite tener un control exacto de lo que haces con el lenguaje R. Así puedes tener control en:

- variables
- gráficas
- de análisis

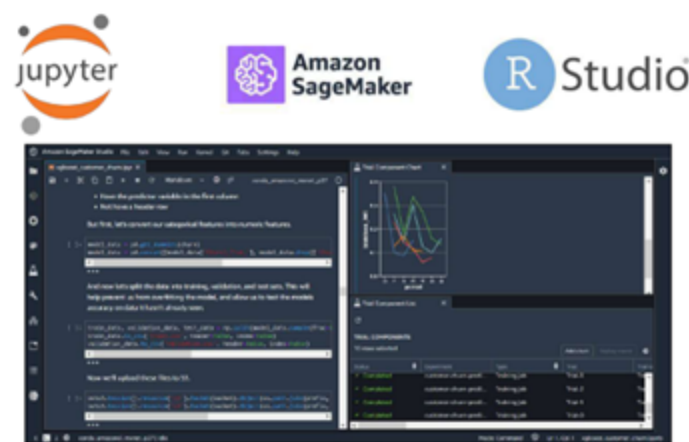
## Azure Machine Learning

Es una herramienta enfocada al machine learning, también está basado en notebooks, te permite también hacer **pipelines** para experimentos o proyectos.

## Amazon Sage Marker

Integro **Jupyter** y **RStudio** para que puedas crear lo que tu necesites, además de ser tan potentes solo te cobran por el uso que les des. Así si en algún momento necesitamos demasiados recursos, los puedes obtener. Pero si en otro momento solo necesitas un poco de recursos, también puedes solicitarlo.

En general pagas por lo que necesites y ocupes.



## Deepnote

Últimamente se volvió muy popular y la razón es que te permite colaborar de manera muy sencilla. Puedes copiar notebooks de manera fácil y sencilla, además de poder desarrollar todo en línea.

# Clase 1: Conociendo nuestros datos de pingüinos

Antes de empezar creare un nuevo entorno para este curso especialmente y así no interferir con las librerías que tengo instaladas,para ello ejecutare el siguiente código en la terminal:

```
#Para el caso de usar conda
conda create --name explorato
```

```
#También se puede usar mamba
mamba create --name explorato
Procederemos a continuar con el curso
```

Primero ejecutaremos, todo esto tenemos que ejecutarlo en la terminal:

```
!pip install --upgrade pip
!pip install palmerpenguins==0.1.4 numpy==1.23.4 pandas==1.5.1 seaborn==0.12.1 matplotlib==3.6.0 empiricaldist==0.6.7 statsmodels==0.13.5 scikit-learn==1.1.2 pyjanitor==0.23.1 session-info
```

```
In [ ]: #Actualizando Los entornos
!pip install --upgrade pip
!pip install palmerpenguins==0.1.4 numpy==1.23.4 pandas==1.5.1 seaborn==0.12.1 matplotlib==3.6.0 empiricaldist==0.6.7 statsmodels=
```

The kernel failed to start as the Python Environment 'explorato' is no longer available. Consider selecting another kernel or refreshing the list of Python Environments.

La anterior celda me falló así que tuve que volver a crear el entorno desde cero con los siguientes comandos en la terminal

```
mamba create --name explorato numpy==1.23.4
```

```
#Luego ejecute
mamba install python==3.10.12
```

```
#Aun que es preferible hacer primero esto y luego
#instalar todos los anteriores paquetes
mamba create --name explorato python==3.10.12
```

```
In [ ]: # Importando Librerías
import empiricaldist
import janitor
import matplotlib.pyplot as plt
import numpy as np
import palmerpenguins
import pandas as pd
import scipy.stats
import seaborn as sns
import sklearn.metrics
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.stats as ss
import session_info
```

## Establecer apariencia general de las gráficas

```
In [ ]: %matplotlib inline
sns.set_style(style='whitegrid')
sns.set_context(context='notebook')
plt.rcParams['figure.figsize'] = (11, 9.4)

penguin_color = {
    'Adelie': '#ff6602ff',
    'Gentoo': '#0f7175ff',
    'Chinstrap': '#c65dc9ff'
}
```

Con el proceso anterior termina la configuración inicial para el espacio de trabajo. En esta clase aprendimos sobre los diferentes tipos de herramientas que hay.

- [Archivo Deepnote](#)