

AL MAL TIEMPO, BUENA DATA

Limpieza de datos con Python

Python Bucaramanga

AL MAL TIEMPO, BUENA DATA



Limpieza de datos con Python



Al mal tiempo, buena data Jun 8 · 8 min read



Sandy Atencio Hernandez

Systems Engineer | Data scientist in progress

Following

1 Follower

About

Jul 5

Limpieza de datos con Python (Parte 2)

En días pasados realice una colaboración con [Al mal tiempo, buena data](#), en donde hablamos sobre Limpieza de datos con Python, abordamos temas sobre la metodología CRISP-DM y como podríamos implementarlo a la hora de trabajar con Python. En este [enlace](#) podrás encontrar la primera parte de limpieza de datos con Python y de ese modo no te sentirás un poco perdido en esta segunda parte 🤖.



AGENDA

1

Qué es ciencia
de datos

2

Herramientas
para ciencia de
datos

3

Metodología
CRISP-DM

4

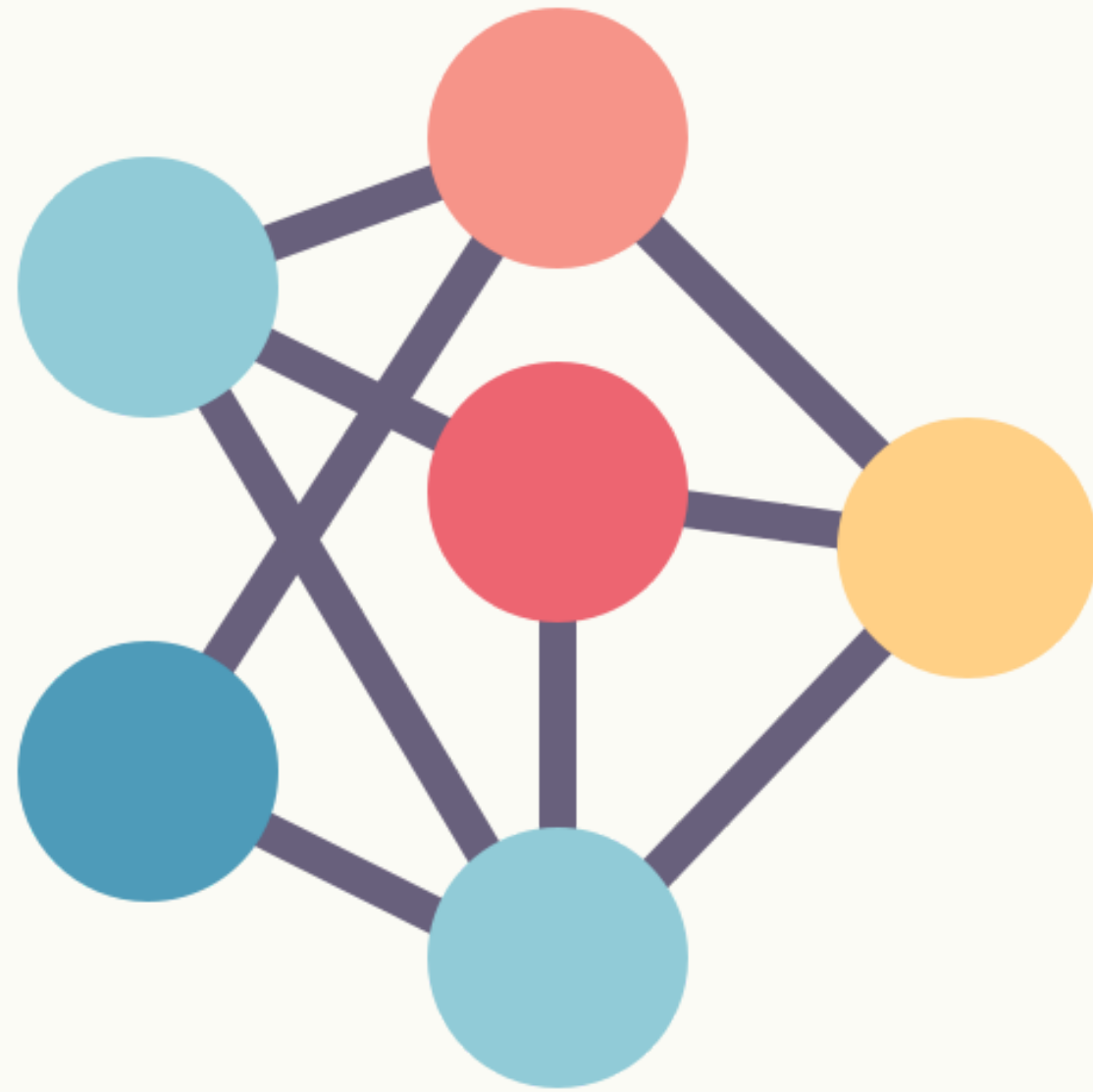
Qué es limpieza
de datos

5

Práctica con
Python

6

Cierre



CIENCIA DE DATOS

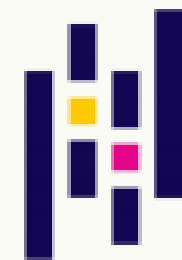
Conceptos, herramientas y tecnologías enfocadas a la extracción, análisis y presentación de información, donde se **genera valor mediante el uso de los datos.**

HERRAMIENTAS

Procesamiento



NumPy



pandas

Visualización

matplotlib



seaborn



plotly

HERRAMIENTAS

IA

 PyTorch

 scikit
learn

 TensorFlow

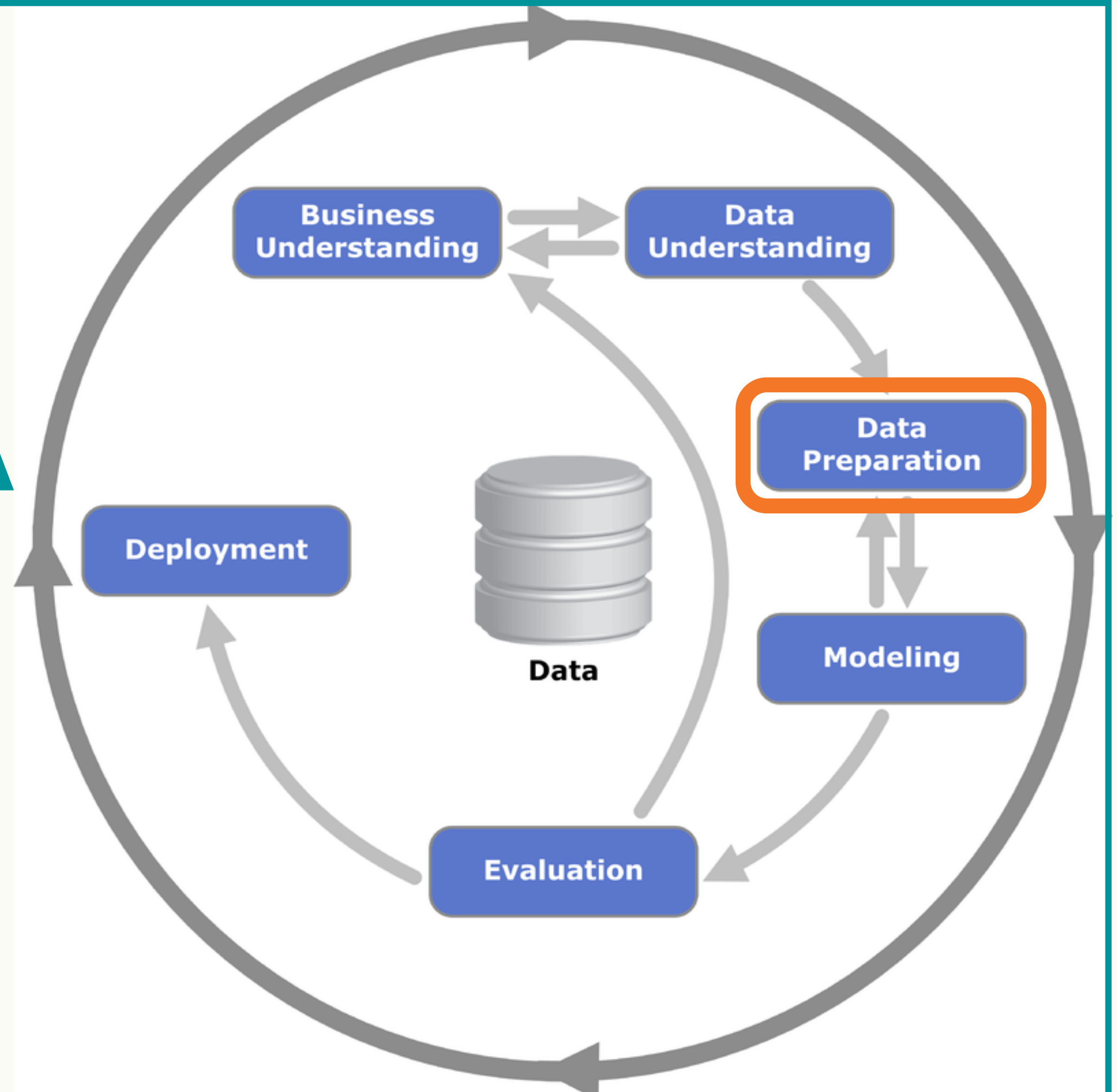
Web Scraping

BeautifulSoup



Scrapy

METODOLOGÍA CRISP-DM





"This is not what I meant when I said 'we need better data cleansing!'"

LIMPIEZA DE DATOS

Tipos de
datos,
rangos y
duplicados

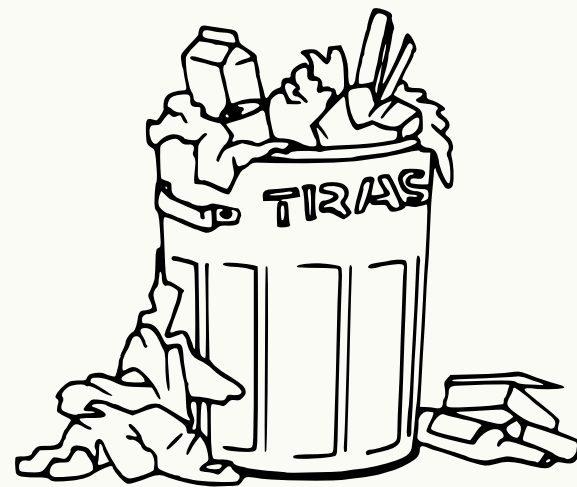
Problemas
de datos
numéricos y
categóricos

Uniformidad
y
completitud

Unión de
varias
fuentes
de datos

¿POR QUÉ ES IMPORTANTE?

Entra basura



Sale basura



ESPACIO DE TRABAJO



Jupyter Notebook




Google Colab




MANOS A LA OBRA!

https://github.com/lauralpezb/Meetup_Python_Bucaramanga


kaggle™




- Home
- Competitions
- Datasets
- Code
- Discussions
- Courses**
- More
 - User Rankings
 - Tags
- Blog
- Documentation
- Progression
- Host a Competition
- View Active Events

**Pandas**


Solve short hands-on challenges to perfect your data manipulation skills.

**Data Visualization**


Make great data visualizations. A great way to see the power of coding!

**Feature Engineering**


Better features make better models. Discover how to get the most out of your data.

**Data Cleaning**

Master efficient workflows for cleaning real-world, messy data.

**Intro to SQL**

Learn SQL for working with databases, using Google BigQuery.

**Advanced SQL**

Take your SQL skills to the next level.

GRACIAS!

@ lauralpezb

