***Let’s start by exploring the two principal concepts of Pandas. Data management and Data analysis.***

Pandas is a powerful Python library that plays a crucial role in data management. It is crucial in business intelligence and data-driven decision-making. Some of the features that Pandas has regarding Data Management are:

**Data Handling:** Pandas provides an easy-to-use structure called Dataframe, which we are going to talk about later. Dfs allow for efficient storage and retrieval of data using a tabular form. This structure is flexible making it easier to organize and access information***. Basically, it’s an excel table with steroids***.

**Next, it’s Data Cleaning:** in the real world, Data is often messy and incomplete ***-You can hear me complain about it very often-*** Pandas offers a suite of tools to address issues like missing data, outliers, and inconsistencies. This ensures that the data used for analysis is accurate and reliable.

**Data Transformation:** Pandas facilitates the transformation of data through operations like grouping, merging, and aggregating. ***-SQL comment-*** This is essential when needing to reshape data to gain insights or prepare it for specific analyses.

***Which leads me to our next concept which is \*NEXT SLIDE\****

**Data Analysis. It** is a core function for people aiming to make informed decisions. In this regards Pandas offers the following capabilities.

**Descriptive Statistics:** Pandas provides statistics functions that help to understand the characteristics of the data. Examples like mean, median, standard deviation, and more.

**Exploratory Data Analysis (EDA):** Pandas supports EDA through its ability to create visualizations and perform data profiling. You can use Pandas in combination with Matplotlib to generate meaningful plots and charts for a comprehensive understanding of the data. ***WE WILL SEE THIS IN MORE DETAIL IN THE LIVE DEMO***

**Reporting:** Pandas integrate seamlessly with other data analysis and visualization tools, enhancing its utility in reporting.

Now let’s talk about the crown jewel of Pandas

A DataFrame is a two-dimensional table, similar to a spreadsheet or SQL table. ***BUT AS I SAID BEFORE, WAY BETTER***

Consists of rows and columns, where each row represents a unique record, and each column represents a different variable or feature.

Each column can have a different data type (LIKE integer, float, string) and is labeled, making it easy to reference and manipulat. On the other hand, rows have a unique index to identify them. In this example, if each row It’s a person 0 would be the first one and so one.

LIVE DEMO

Df.head()

Df.tail()

Df.describe()

Df.count()

Df.count(axis=)

dups\_df = raw\_df.drop\_duplicates()

nans\_df = raw\_df.dropna()

raw\_df.fillna(999,inplace=True)

raw\_df.head()

raw\_df.to\_csv('./example.csv')

SELECT A COLUMN

raw\_df['Points']

raw\_df[['FamilyName','Points']]

FILTERING

raw\_df[raw\_df['Points'] == 25]

TRANSFORMING DATA

raw\_df['Winner'] = np.where(raw\_df['Points']>=25,'Yes','No')

raw\_df[raw\_df['Winner'] == 'Yes']