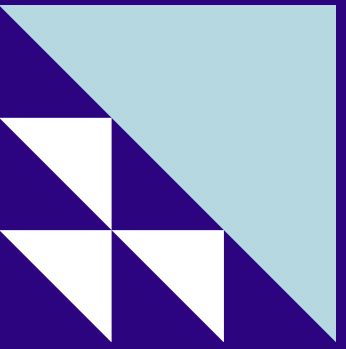


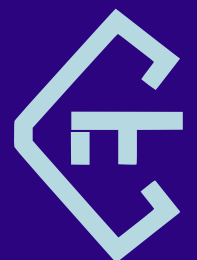


MODULE 1 : Data Exploration



DATA-MANIPULATION-1

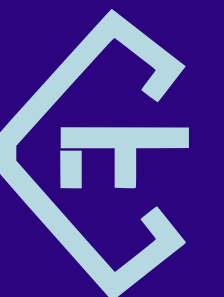
Wrangling Data to extract meaningful insights



Club Informatique & Télécom
Data Cell

Today's Menu

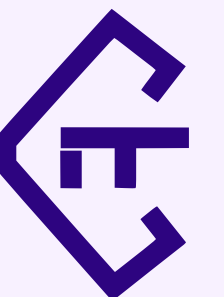
1. What is Data Wrangling ?
2. pandas Library
3. pandas Data Structure
4. Viewing / Inspecting Data
5. Selecting Data
6. Cleaning Data
7. lab

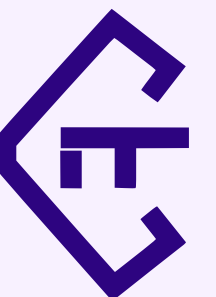


Data Wrangling

Data Wrangling is the process of transforming and structuring data from one raw form into a desired format with the intent of improving data quality and making it more consumable and useful for analytics or machine learning. It's also sometimes called data munging.

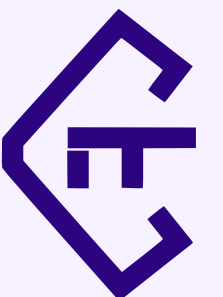
~ [alteryx.com](https://www.alteryx.com)





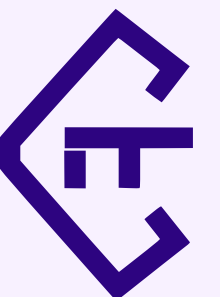
Viewing / Inspecting data

<code>df.head(N)</code>	Displays first N rows. By default N=5
<code>df.tail(N)</code>	Displays last N rows. By default N=5
<code>df.shape</code>	Number of rows and columns
<code>df.info()</code>	Index, DataType, and Memory information
<code>df.describe()</code>	Summary statistics for numerical columns
<code>df['col'].value_counts()</code>	View a summary of the unique values in a Series



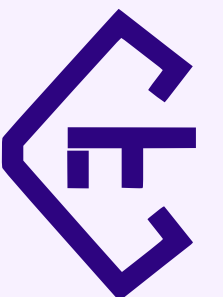
Selecting data

<code>df['col']</code>	Single column by name
<code>df[['col1', 'col2']]</code>	Multiple columns by name.
<code>df.iloc[row_idx, col_idx]</code>	Select rows and columns by index
<code>df.loc[row_idx, col_label]</code>	Select rows by index and columns by label



Cleaning data

<code>pd.isnull(df)</code>	Check for null values, returns a Boolean array
<code>df.dropna(axis=0)</code>	<ul style="list-style-type: none">• By default (axis=0), drops all rows that contain NULL values• axis=1, drops all columns that contain NULL values
<code>df.fillna(value)</code>	Replace all NULL values with the specified value
<code>df['col'].replace([1,2], ['one','two'])</code>	In column 'col', replace all 1 with 'one' and 2 with 'two'
<code>df.rename(columns={'old': 'new'})</code>	Rename column 'old' to 'new'



Ready, Set, Code!

