# Data manipulation and visualization cheat sheet

## Working with columns: pandas

|  |  |
| --- | --- |
| **Task** | **Operation** |
| Selecting a column | df[['a', 'b', 'c']] |
| Dropping a column | df.drop('a', axis='columns’) |
| Calculating a column | df['new\_column'] = df['x'] + df['y'] |
| Renaming a column | df.rename(columns = {'new\_name':'old\_name'}) |

## Working with rows: pandas

|  |  |
| --- | --- |
| **Task** | **Operation** |
| Sorting rows | df.sort\_values(by=['col'], ascending=[False]) |
| Filtering rows | df[df['col'] == 'x'] |

Aggregating and summarizing: pandas

Graphical user interface, application

Description automatically generated

Aggregating and summarizing: pandas

df\_a.merge(df\_b)

*Default merge type = inner*

df\_a.merge(df\_b, how='left’)

*Left outer join*

Chart, bar chart

Description automatically generated

Aggregating and summarizing: pandas

|  |  |
| --- | --- |
| **Argument** | **What it does** |
| data | What dataset to use |
| x | What variable goes on x axis? |
| y | What variable goes on y axis? |
| color, bins, etc. | Other ways to customize the plot |

Aggregating and summarizing: pandas

|  |  |
| --- | --- |
| **Type of plot** | **Function** |
| Countplot/bar chart | sns.countplot() |
| Histogram | sns.displot() |
| Boxplot | sns.boxplot() |
| Scatterplot | sns.boxplot() |
| Pairplot | sns.pairplot() |

Aggregating and summarizing: pandas

|  |  |
| --- | --- |
| **Type of label** | **Function** |
| X axis | plt.xlabel() |
| Y axis | plt.ylabel() |
| Title | plt.title() |