Python Learning

* Data Munging Basic

→ Filter and select data
- Using pandas and Numpy
  ↓
  DataFrame, Series
- Retrieve data

→ Treat missing value
- Check what is missing → df.isnull()
- Fill the missing value → df.fillna(0) (fill 0)
  - ↳ df.fillna({0: 0.1, 5: 1.25}) (fill specific cols)
  - ↳ df.fillna(method='ffill') (forward fill)
- count missing value.
  df.isnull().sum()

→ Filtering out missing value → df.dropna()
↳ df.dropna(how='all') → drop if all missing.

→ Remove duplicates
df.duplicated() → display True for duplicates.
df.drop_duplicates() → Remove duplicates.
df.drop_duplicates(['col_name']) → specific col

→ concatenating data
pd.concat([df1, df2], axis=1)
↳ optional

→ dropping data
pd.Drop([0, 2], axis=1)

→ adding data
pd.join(df1, df2.col_name)
pd.append

→ sort data
df.sort_values(by=[5], ascending=[False])

→ para grouping and aggregating
df.groupby([df.colname])