

LECTURE 5

# More on Pandas

Pivot, GroupBy, Merge

**Sean Kang**

# Using Pivot

# Using Pivot in Pandas to aggregate

- You can calculate the counts on a specific numerical column
  - But you might get some un-expected NaN
  - Look at the baby names and aggregating by Sex across names.
- Debugging in VS Code with jupyter lab files?
  - Yes: with print statements
- Can I use something else better for real time debugging?
  - Yes: You can use other python IDE such as PyCharm

# Using GroupBy

- We can do the similar operations as Pivot using GroupBy
  - We need to be careful as well here.
- We need to specify which columns to focus on, otherwise python gets confused
  - Esp with assuming data types
- Need to evaluate the data types, at each step of the data break down.
  - Compare the columns in `bn_agg_df.columns[[0]].tolist()`
  - Compared to `df_result = grouped['Sales'].agg([showcount])`

# Merge of Dataframe

- Two different ways to do it
  - `<dataframe>.merge(dataframe2)`
  - `pd.merge()`
- Left, right, inner, outer, cross

# Plotting with Dataframe

- Two different ways to do it
  - `<dataframe>.plot` – part of pandas
  - `plt.plot()` – using matplotlib.
- Left, right, inner, outer, cross