C5

Getting Started with pandas

CHAPTER ASSIGNMENT

Name _____ Class _____ No. _____ Date _____

> Answer every question in the space given. Show your working. You may use the chapter page to help you.

Questions from the Guru Jazzy chapter page (based on *Python for Data Analysis*, 3rd edition, Wes McKinney, O'Reilly).

5.1 Series

1. Make a Series of five classmates' heights with their names as the index. Select everyone taller than the mean.

2. Make a second Series of weights for four of the same people plus one new person, in a different order. Add the two Series together and explain every NaN.

5.2 DataFrame and Index Objects

1. Build a DataFrame of five books with title, author, year and pages. Add a column "long" that is True when pages > 300.

2. Print frame.T for your books table. When would a transposed table be easier to read?

5.3 Reindexing and Dropping

1. Reindex the books table so its rows come out in reverse order, then drop the "pages" column. Check the original still has it.

2. Make pd.Series(["blue", "purple", "yellow"], index=[0, 2, 4]) and reindex it to range(6) with and without method="ffill".

5.4 Selecting with loc and iloc

1. Use the selector below. For each expression, predict which cells light up and whether the result is a Series, a DataFrame or a single value.

```
data = pd.DataFrame(np.arange(16).reshape((4, 4)),
                    index=["Ohio", "Colorado", "Utah", "New York"],
                    columns=["one", "two", "three", "four"])

data["two"]
data[["three", "one"]]
data.loc["Colorado", ["two", "three"]]
data.iloc[2]
data.iloc[[2, 1], [3, 0, 1]]
data.loc[:"Utah", "two"]
data.iloc[:2, 1:3]
data.loc[data["three"] > 5]
data.loc["Utah", "four"]
```

2. Write two ways to get the "four" column for Colorado and Utah: once with loc, once with iloc.

3. Try the three error buttons. Explain in one sentence each why they fail.

```
data = pd.DataFrame(np.arange(16).reshape((4, 4)),
                    index=["Ohio", "Colorado", "Utah", "New York"],
                    columns=["one", "two", "three", "four"])

data[1]
data.loc["Texas"]
data.iloc[4]
```

5.5 Arithmetic, Functions and Sorting

1. Make two Series of weekly sales for two shops with partly different product names. Add them with + and with add(fill_value=0). Which answer is right for "total sales"?

2. On the books table, use apply to get max − min of every numeric column, then sort the books by year, newest first.

5.6 Descriptive Statistics

1. Run describe() on the books table. Which numbers are useful for the "year" column, and which are meaningless?

2. Make a Series of 20 favourite colours from your class. Find the top three with value_counts and filter to only red or blue with isin.

End of assignment · Check your answers against the chapter at gurujazzy.com/subjects/books/python-for-data-analysis/chapter-5-getting-started-with-pandas/