# Merge pandas

Trinity, Anjali, Sara, and Treena

# Follow the Way of the Panda:

1. #FIRST, make sure that you have python on your machine… YOU NEED PYTHON INSTALLED:

#instructions for macos: https://www.dataquest.io/blog/installing-python-on-mac/
#instructions for windows: https://www.dataquest.io/blog/installing-python-on-windows/
#python download page: https://www.python.org/downloads/

2. #Change to working branch using:

`git checkout <name of branch>`

3. #Create a Virtual Environment Using(if you already have one in mind skip to 4):

`python3 -m <name of your virtual environment> .venv`

4. #activate the virtual environment

`source .venv/bin/activate`

#Go to next slide if the csvs you want to merge are already inside your current branch
Step 5. #open python shell in terminal using:
```
python3
python
```
 #depening on the version you have

Step 6. #import pandas:
```
Import pandas as pd
```
*to install(if you don't have it already):
```
install pandas
```
#if this does not work this might mean you are not in the shell refer back to step 5
 #to check if pandas is installed/check version:
```
pd.  version
```

Step 7. #have pandas read your csv file, make your changes, (add columns and so forth…)
and then print file in terminal:
```
df = pd.read_csv('<name of csv>')
```
#print csv
```
df
```
# Save the altered csv data to a new CSV file
```
df.to_csv("<name of updated csv>.csv" , index=False)
```
Step 8. #commit and push to team branch
```
git commit -m "my csv to my branch"
git push origin/<name of branch for the merge>
```

Step 8. #Change to the team branch with the csv files you want to combine
```
git checkout <name of branch for the merge>
```

# Shy Panda (Working inside my shell)

# Pandas like to be together (merge)

```python
#in this branch open shell, again
python3
import pandas as pd

# Read the first CSV file
df1 = pd.read_csv("file1.csv")

# Read the second CSV file
df2 = pd.read_csv("file2.csv")

# Merge the two dataframes based on the common columns and rows(it's easiest to make
sure the csv files have the same column names —>(df.rename(columns={ 'Name': 'name',
'Food': 'food'}, inplace=True)

merged_df = pd.merge(df1, df2, on=["name", "food", "animal", "fact"], how="outer")

#take a look!
merged_df

# Save the merged data to a new CSV file
merged_df.to_csv("merged.csv", index=False)

#yay
```
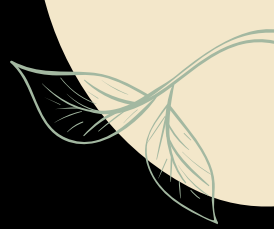
# Oh Yeah, Give Me Paw!

- First, we worked on individual branches, then pushed everything to 'Team1' branch, then we combined the csv's two at a time!

- Then we pushed the final csv file to the Main branch!

# Watch out! Cowapanda! (Challenges):

- Repeated Cell values/Doubles and extra commas made it difficult to merge the first set of csv files initially which meant we needed an extra step to delete the doubles.
- We had a lot of issues pushing our personal branch to the Team 1 branch