**URL for project assignment:**

<https://www.springboard.com/workshops/data-science-career-track/learn#/curriculum/448>

***Solution for question1:***

import pandas as pd

import json

import operator

from pandas.io.json import json\_normalize

sample\_json\_df = pd.read\_json(r'C:\Users\anshu\Documents\assign\data\_wrangling\_json\data\world\_bank\_projects.json')

# Initialize an empty dictionary: counts\_dict

counts\_cntry = {}

#looping through dataframe

for column in sample\_json\_df['countryshortname']:

if column in counts\_cntry.keys():

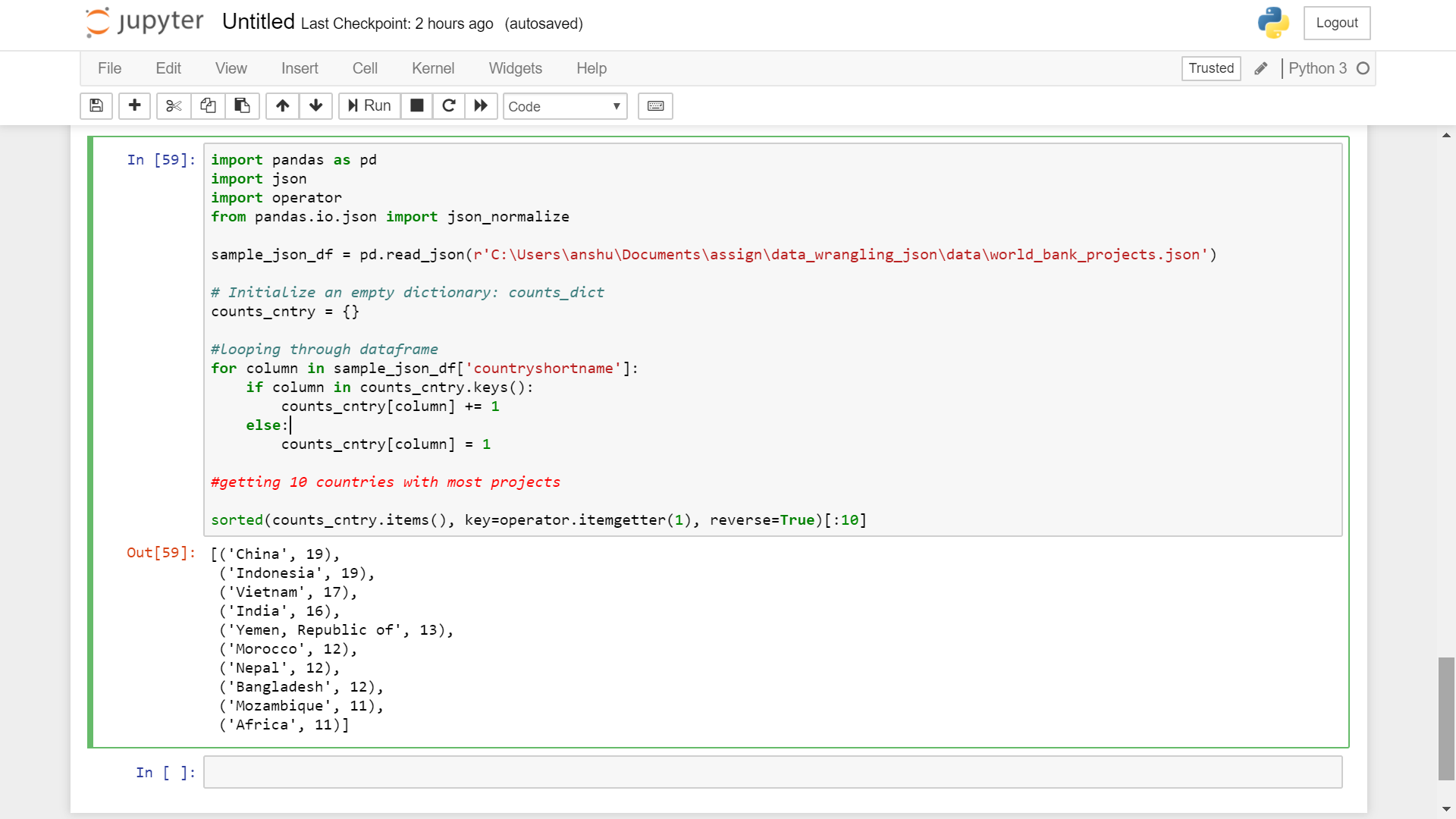
counts\_cntry[column] += 1

else:

counts\_cntry[column] = 1

#getting 10 countries with most projects

sorted(counts\_cntry.items(), key=operator.itemgetter(1), reverse=True)[:10]



***Solution for question2:***

import pandas as pd

import json

import operator

from pandas.io.json import json\_normalize

df = pd.read\_json(r'C:\Users\anshu\Documents\assign\data\_wrangling\_json\data\world\_bank\_projects.json')

# also load in the raw json to wrangle the nested fields later

with open(r'C:\Users\anshu\Documents\assign\data\_wrangling\_json\data\world\_bank\_projects.json') as f:

raw = json.load(f)

df\_themes = json\_normalize(raw, 'mjtheme\_namecode', ['id'])

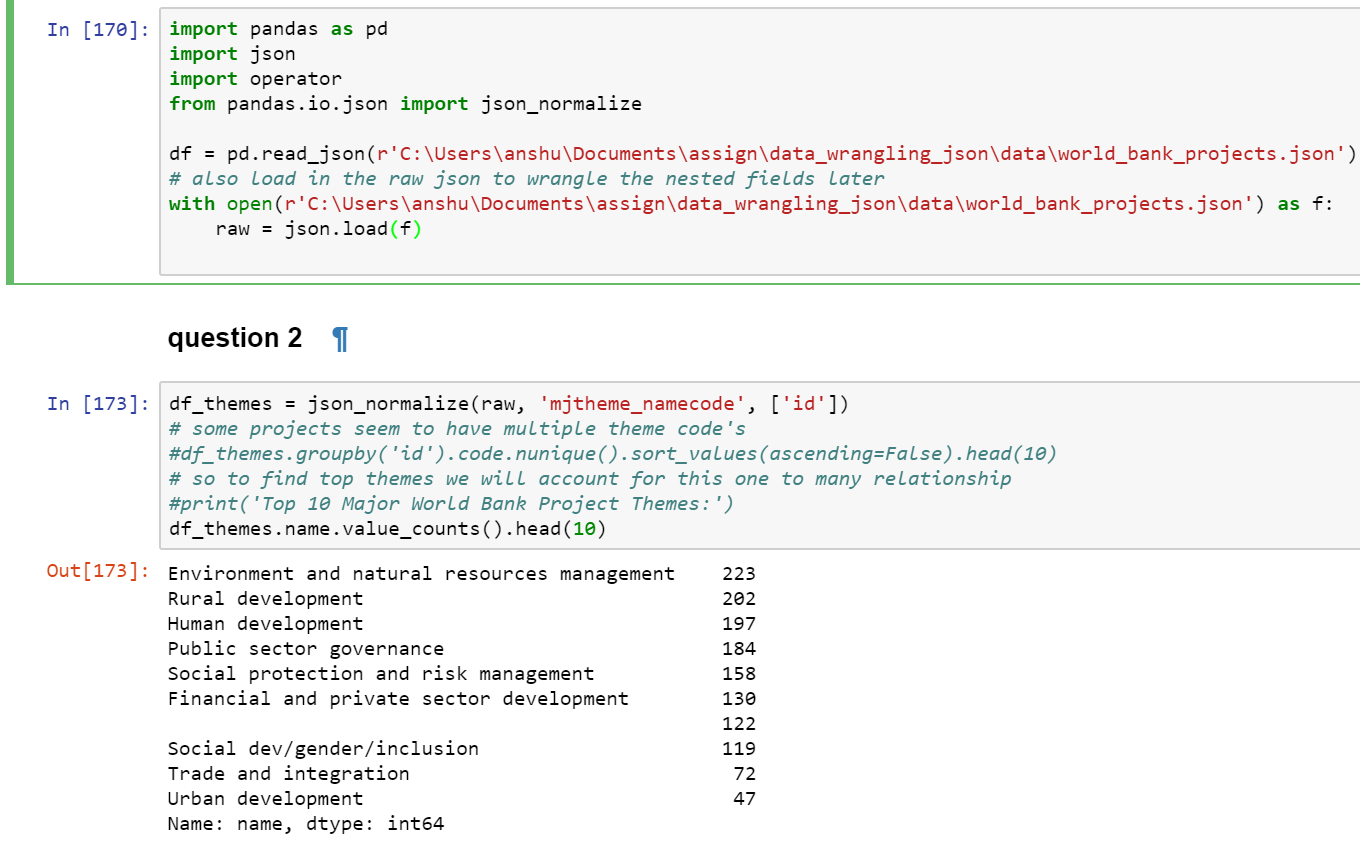
# some projects seem to have multiple theme code's

#df\_themes.groupby('id').code.nunique().sort\_values(ascending=False).head(10)

# so to find top themes we will account for this one to many relationship

#print('Top 10 Major World Bank Project Themes:')

df\_themes.name.value\_counts().head(10)



***Solution for question3:***

df\_themes\_name\_to\_code = df\_themes.groupby('name').code.max().sort\_values(ascending=False)

df\_themes\_name\_to\_code = df\_themes\_name\_to\_code[df\_themes\_name\_to\_code.index != '']

df\_themes\_name\_to\_code = pd.DataFrame(df\_themes\_name\_to\_code,columns=['code'])

df\_themes\_name\_to\_code['name\_clean'] = df\_themes\_name\_to\_code.index

df\_themes\_code\_to\_name = df\_themes\_name\_to\_code.set\_index(['code'])

df\_themes\_clean = df\_themes.merge(df\_themes\_code\_to\_name,how='outer',left\_on=['code'],right\_index=True)

df\_themes\_clean

