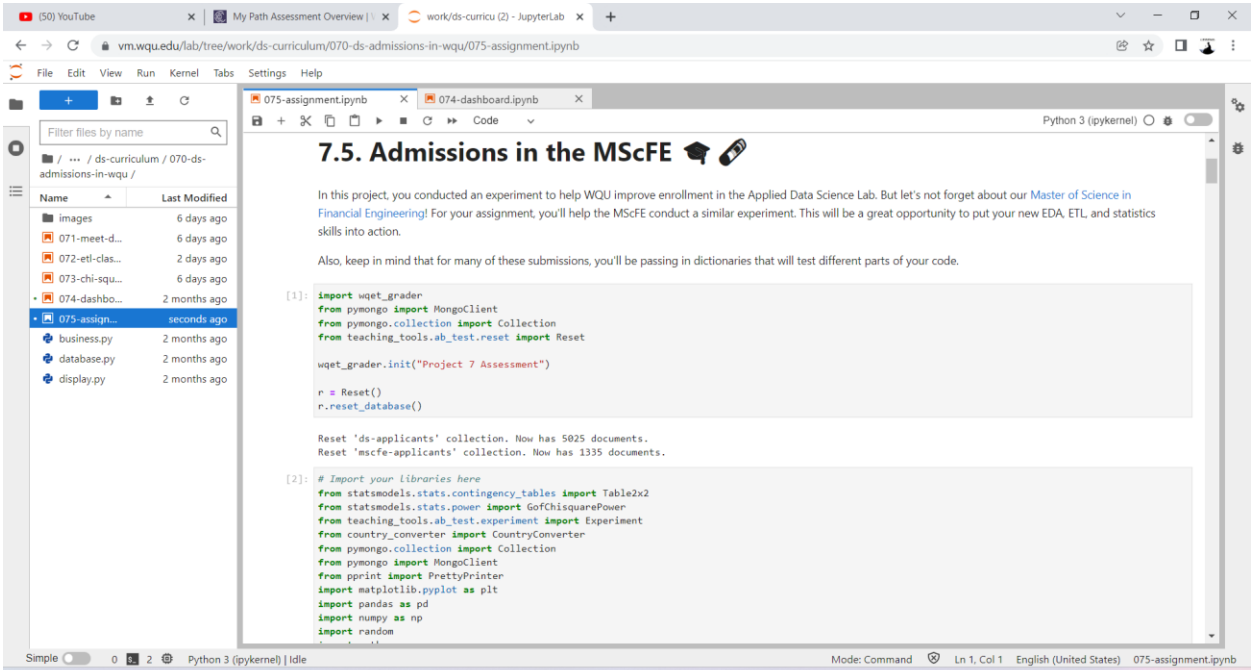
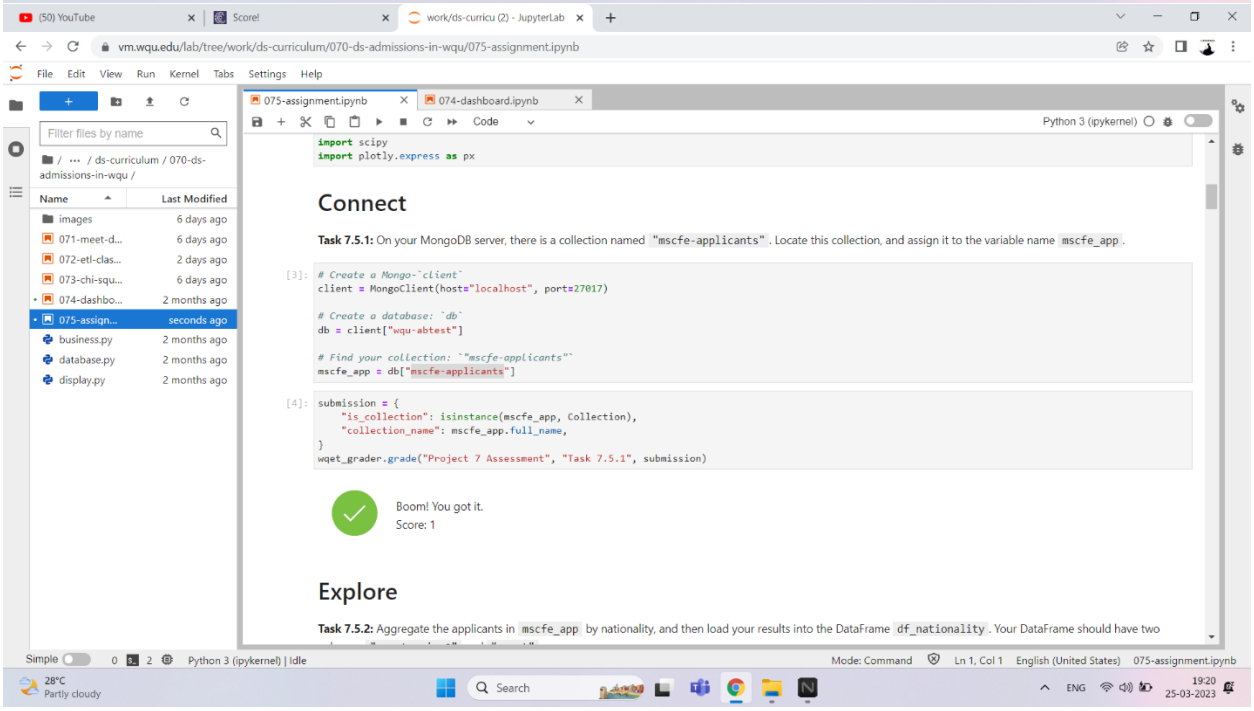
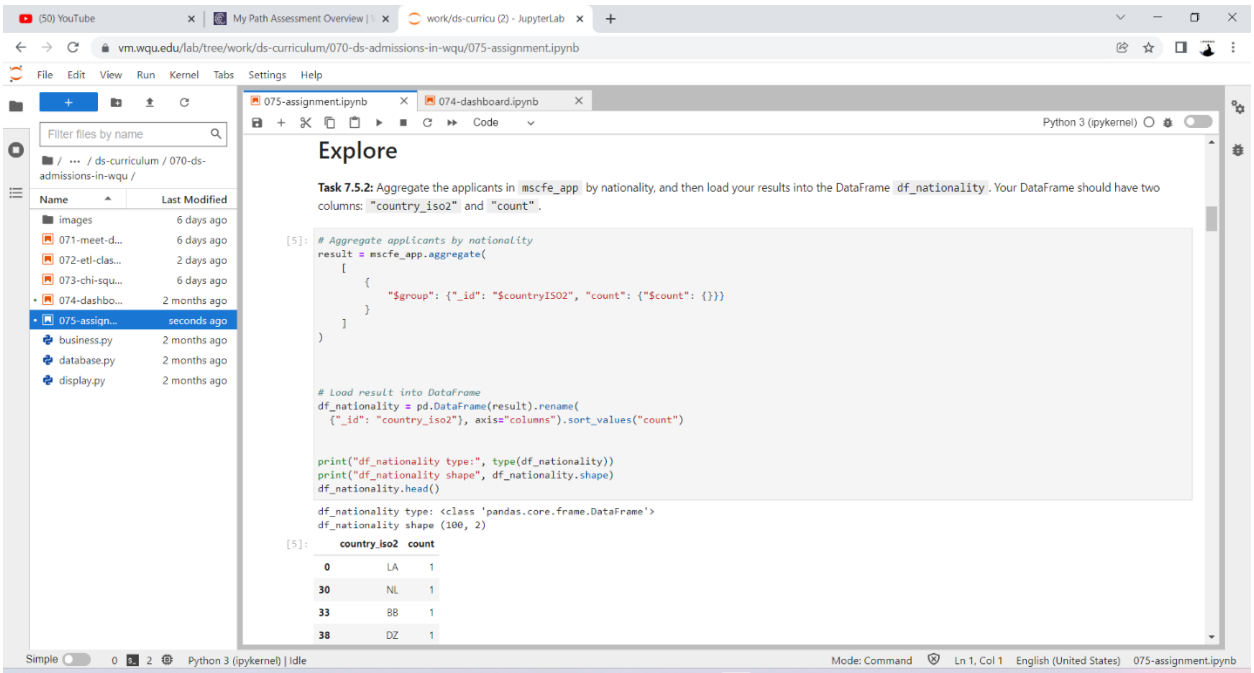
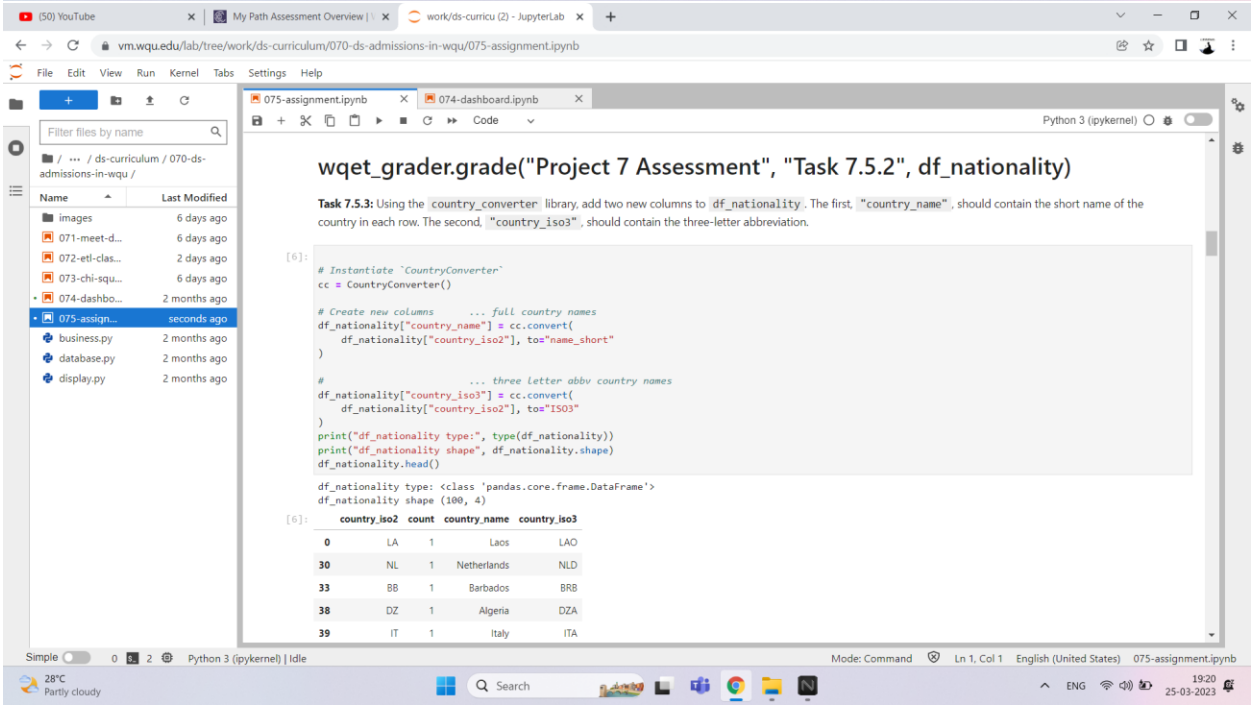
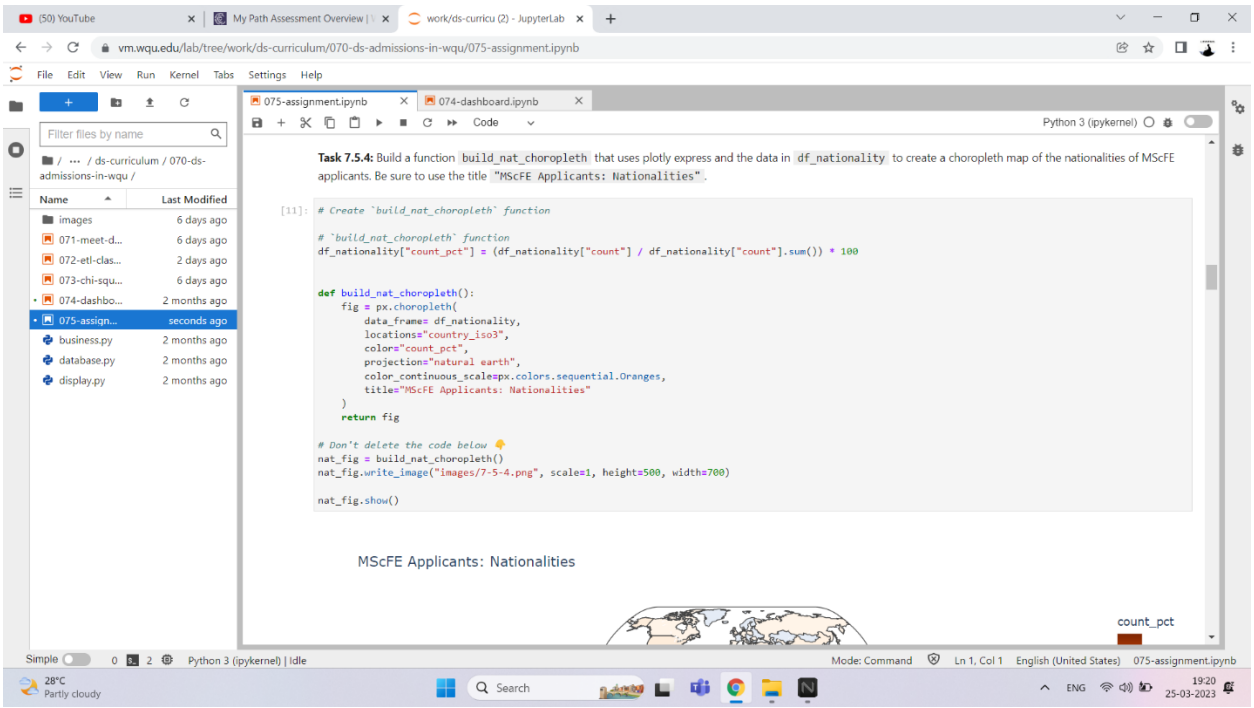
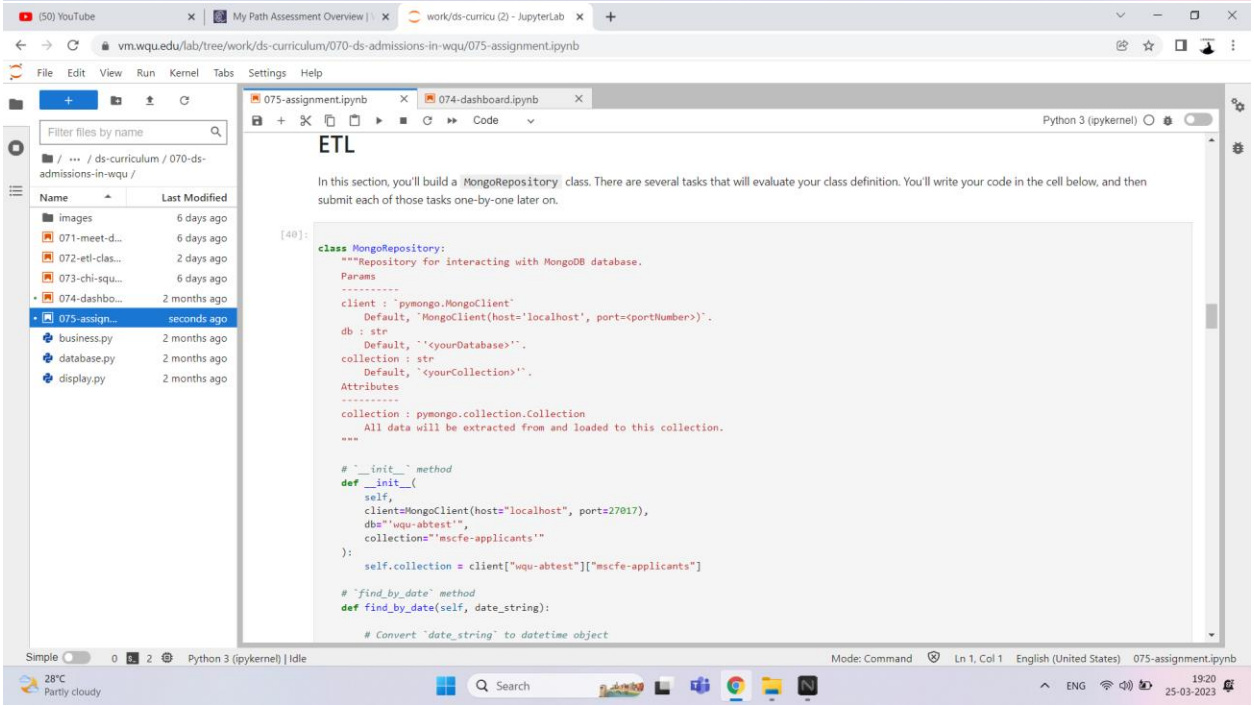

(50) YouTube
My Path Assessment Overview
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
7.5. Admissions in the MScFE
In this project, you conducted an experiment to help WQU improve enrollment in the Applied Data Science Lab. But let's not forget about our Master of Science in Financial Engineering! For your assignment, you'll help the MScFE conduct a similar experiment. This will be a great opportunity to put your new EDA, ETL, and statistics skills into action.
Also, keep in mind that for many of these submissions, you'll be passing in dictionaries that will test different parts of your code.
[1]:
import wqet_grader
from pymongo import MongoClient
from pymongo.collection import Collection
from teaching_tools.ab_test.reset import Reset
wqet_grader.init("Project 7 Assessment")
r = Reset()
r.reset_database()
Reset 'ds-applicants' collection. Now has 5025 documents.
Reset 'mscfe-applicants' collection. Now has 1335 documents.
[2]:
# Import your libraries here
from statsmodels.stats.contingency_tables import Table2x2
from statsmodels.stats.power import GofChisquarePower
from teaching_tools.ab_test.experiment import Experiment
from country_converter import CountryConverter
from pymongo.collection import Collection
from pymongo import MongoClient
from pprint import PrettyPrinter
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import random
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:19 25-03-2023

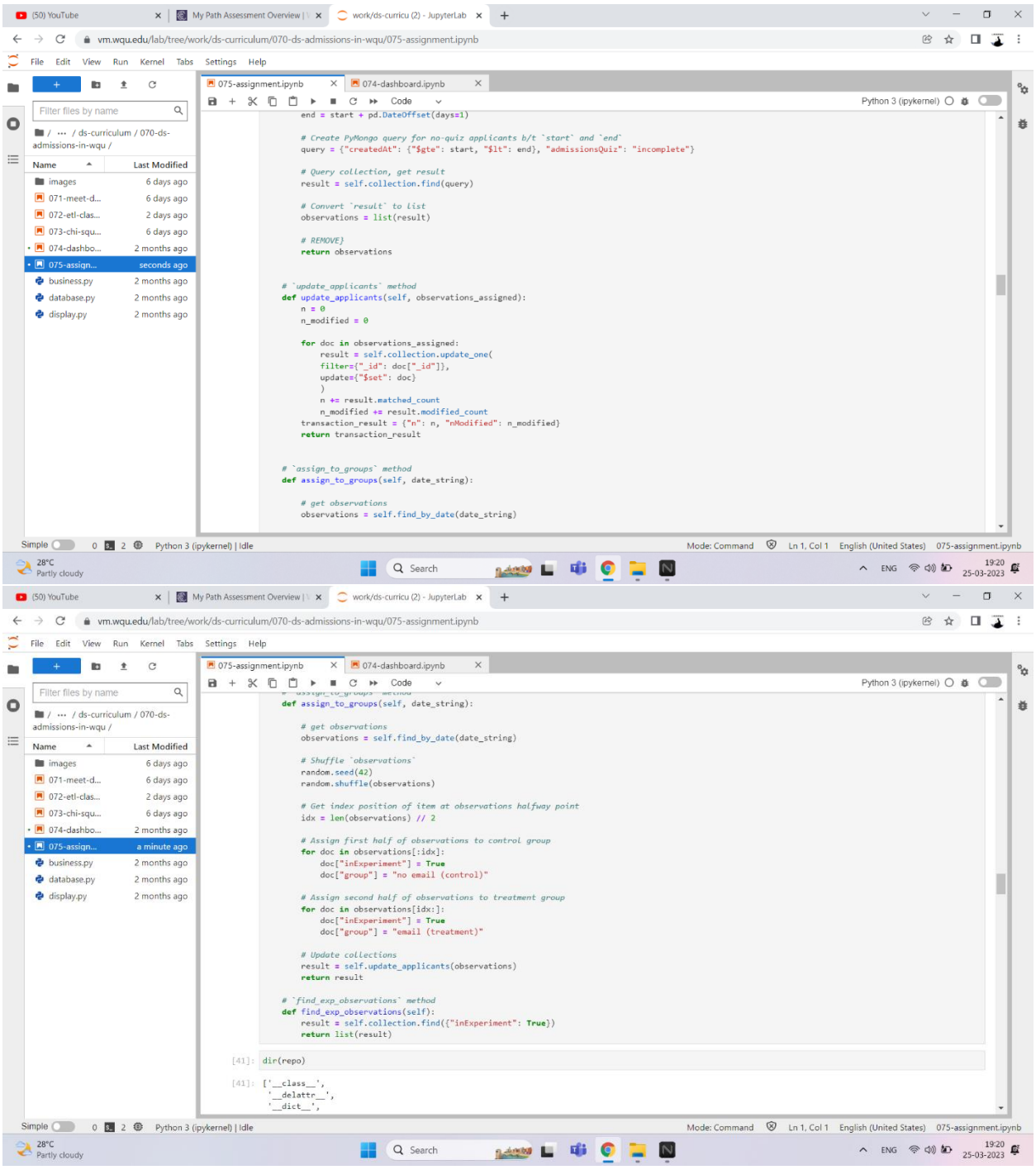

(50) YouTube
Score!
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
import scipy
import plotly.express as px
Connect
Task 7.5.1: On your MongoDB server, there is a collection named "mscfe-applicants". Locate this collection, and assign it to the variable name mscfe_app.
[3]:
# Create a Mongo-`client`
client = MongoClient(host="localhost", port=27017)
# Create a database: `db`
db = client["wqu-abtest"]
# Find your collection: `"mscfe-applicants"`
mscfe_app = db["mscfe-applicants"]
[4]:
submission = {
"is_collection": isinstance(mscfe_app, Collection),
"collection_name": mscfe_app.full_name,
}
wqet_grader.grade("Project 7 Assessment", "Task 7.5.1", submission)
Boom! You got it.
Score: 1
Explore
Task 7.5.2: Aggregate the applicants in mscfe_app by nationality, and then load your results into the DataFrame df_nationality. Your DataFrame should have two
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023


(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Explore
Task 7.5.2: Aggregate the applicants in mscfe_app by nationality, and then load your results into the DataFrame df_nationality. Your DataFrame should have two columns: "country_iso2" and "count".
[5]:
# Aggregate applicants by nationality
result = mscfe_app.aggregate(
[
{
"$group": {"_id": "$countryISO2", "count": {"$sum": {}}}
}
]
)
# Load result into DataFrame
df_nationality = pd.DataFrame(result).rename(
{"_id": "country_iso2"}, axis="columns").sort_values("count")
print("df_nationality type:", type(df_nationality))
print("df_nationality shape", df_nationality.shape)
df_nationality.head()
df_nationality type: <class 'pandas.core.frame.DataFrame'>
df_nationality shape (100, 2)
[5]:
country_iso2 count
0 LA 1
30 NL 1
33 BB 1
38 DZ 1
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023

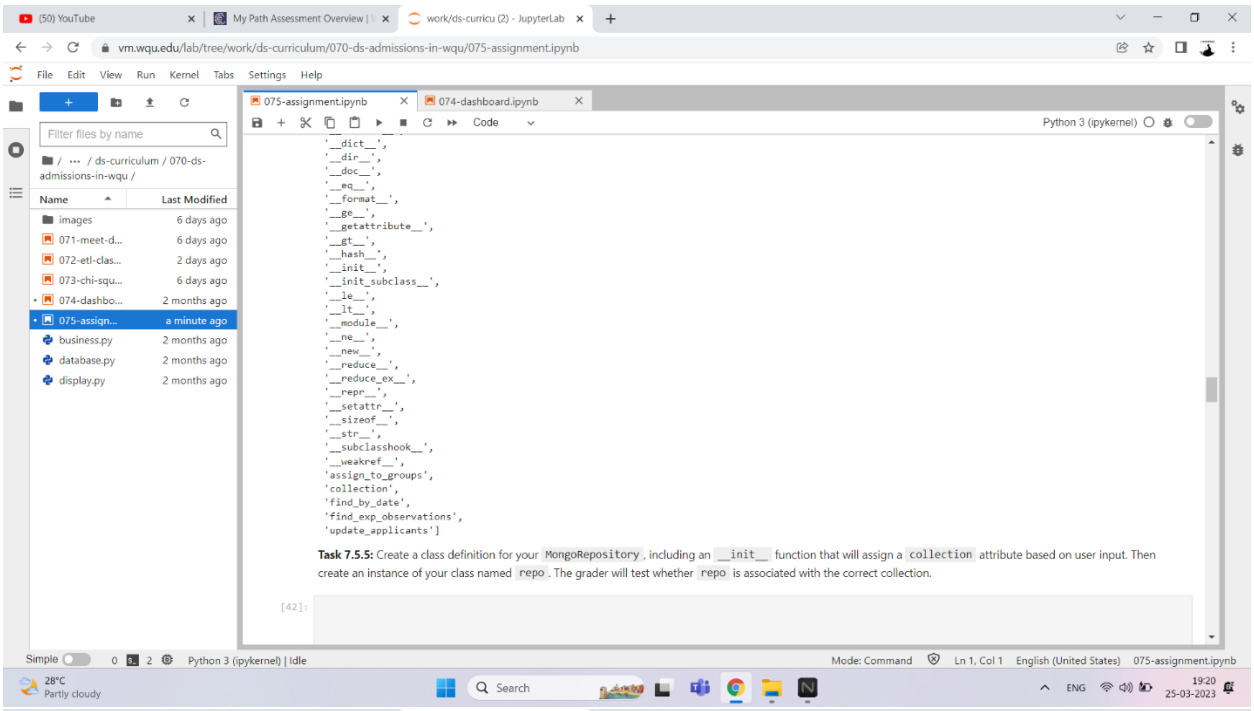

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
wqet_grader.grade("Project 7 Assessment", "Task 7.5.2", df_nationality)
Task 7.5.3: Using the country_converter library, add two new columns to df_nationality. The first, "country_name", should contain the short name of the country in each row. The second, "country_iso3", should contain the three-letter abbreviation.
[6]:
# Instantiate `CountryConverter`
cc = CountryConverter()
# Create new columns ... full country names
df_nationality["country_name"] = cc.convert(
df_nationality["country_iso2"], to="name_short"
)
# ... three letter abbv country names
df_nationality["country_iso3"] = cc.convert(
df_nationality["country_iso2"], to="ISO3"
)
print("df_nationality type:", type(df_nationality))
print("df_nationality shape", df_nationality.shape)
df_nationality.head()
df_nationality type: <class 'pandas.core.frame.DataFrame'>
df_nationality shape (100, 4)
[6]:
country_iso2 count country_name country_iso3
0 LA 1 Laos LAO
30 NL 1 Netherlands NLD
33 BB 1 Barbados BRB
38 DZ 1 Algeria DZA
39 IT 1 Italy ITA
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023

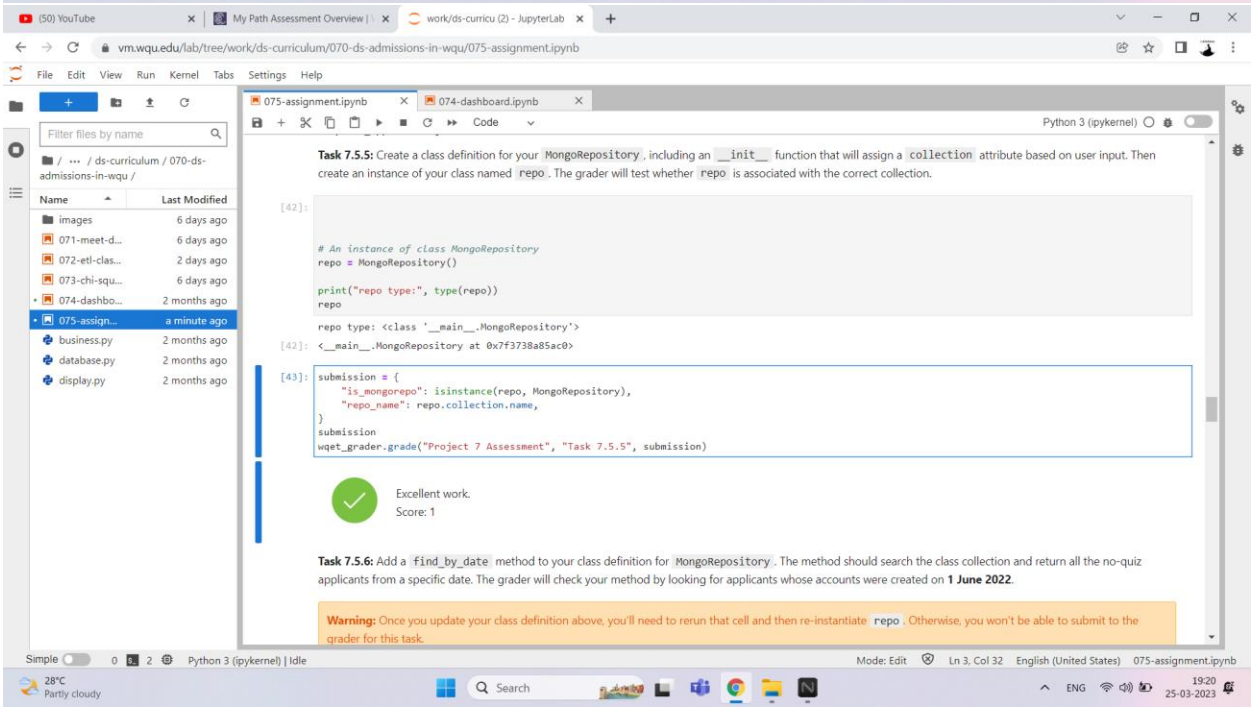

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.4: Build a function build_nat_choropleth that uses plotly express and the data in df_nationality to create a choropleth map of the nationalities of MScFE applicants. Be sure to use the title "MScFE Applicants: Nationalities".
[11]:
# Create `build_nat_choropleth` function
# `build_nat_choropleth` function
df_nationality["count_pct"] = (df_nationality["count"] / df_nationality["count"].sum()) * 100
def build_nat_choropleth():
fig = px.choropleth(
data_frame= df_nationality,
locations="country_iso3",
color="count_pct",
projection="natural earth",
color_continuous_scale=px.colors.sequential.Oranges,
title="MScFE Applicants: Nationalities"
)
return fig
# Don't delete the code below
nat_fig = build_nat_choropleth()
nat_fig.write_image("images/7-5-4.png", scale=1, height=500, width=700)
nat_fig.show()
MScFE Applicants: Nationalities
count_pct
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023

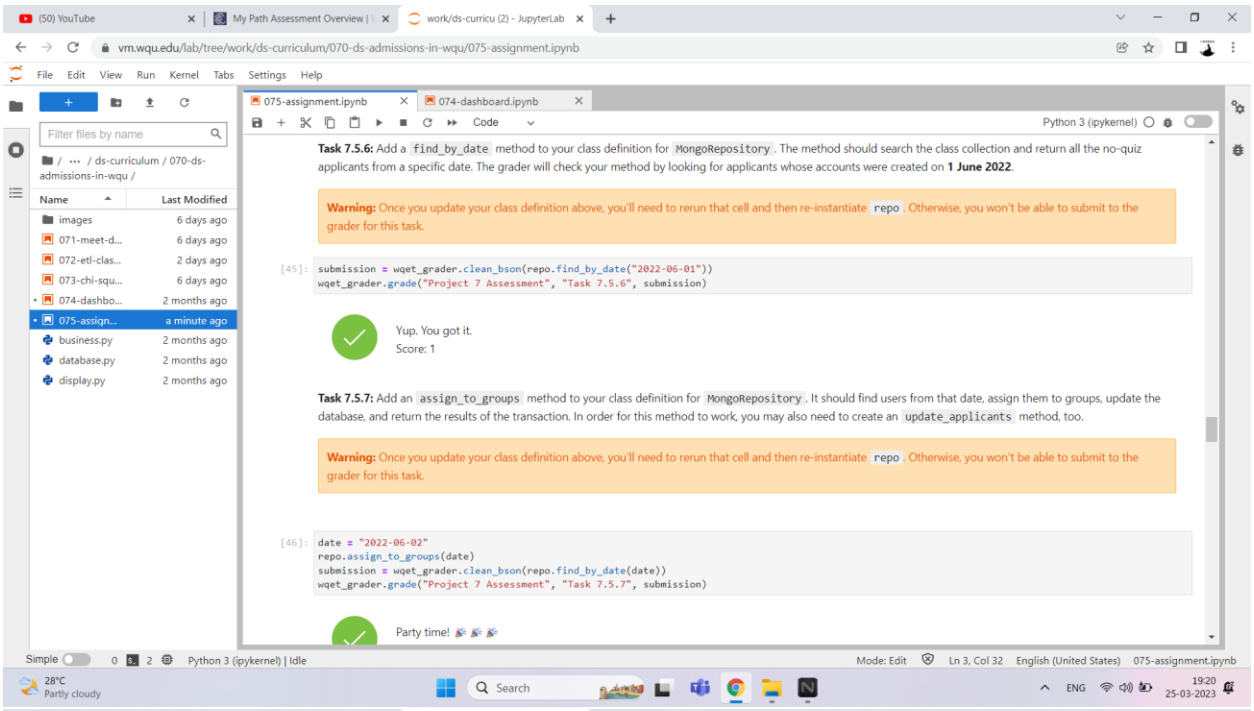

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
ETL
In this section, you'll build a MongoRepository class. There are several tasks that will evaluate your class definition. You'll write your code in the cell below, and then submit each of those tasks one-by-one later on.
[40]:
class MongoRepository:
"""Repository for interacting with MongoDB database.
Params
client : `pymongo.MongoClient`
Default, `MongoClient(host='localhost', port=<portNumber>)`.
db : str
Default, `'<yourDatabase>'`.
collection : str
Default, `<yourCollection>'`.
Attributes
collection : pymongo.collection.Collection
All data will be extracted from and loaded to this collection.
"""
# `__init__` method
def __init__(
self,
client=MongoClient(host="localhost", port=27017),
db="'wqu-abtest'",
collection="'mscfe-applicants'"
):
self.collection = client["wqu-abtest"]["mscfe-applicants"]
# `find_by_date` method
def find_by_date(self, date_string):
# Convert `date_string` to datetime object
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023

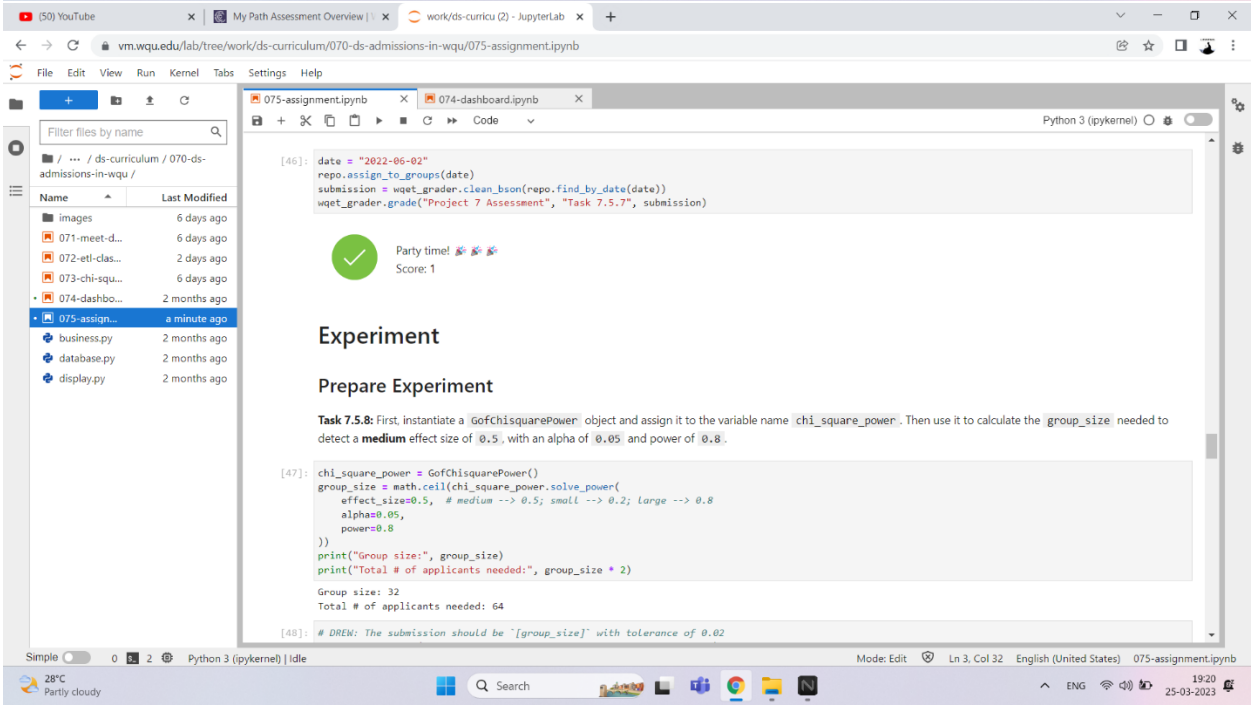

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
end = start + pd.DateOffset(days=1)
# Create PyMongo query for no-quiz applicants b/t `start` and `end`
query = {"createdAt": {"$gte": start, "$lt": end}, "admissionsQuiz": "incomplete"}
# Query collection, get result
result = self.collection.find(query)
# Convert `result` to list
observations = list(result)
# REMOVE}
return observations
# `update_applicants` method
def update_applicants(self, observations_assigned):
n = 0
n_modified = 0
for doc in observations_assigned:
result = self.collection.update_one(
filter={"_id": doc["_id"]},
update={"$set": doc}
)
n += result.matched_count
n_modified += result.modified_count
transaction_result = {"n": n, "nModified": n_modified}
return transaction_result
# `assign_to_groups` method
def assign_to_groups(self, date_string):
# get observations
observations = self.find_by_date(date_string)
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023

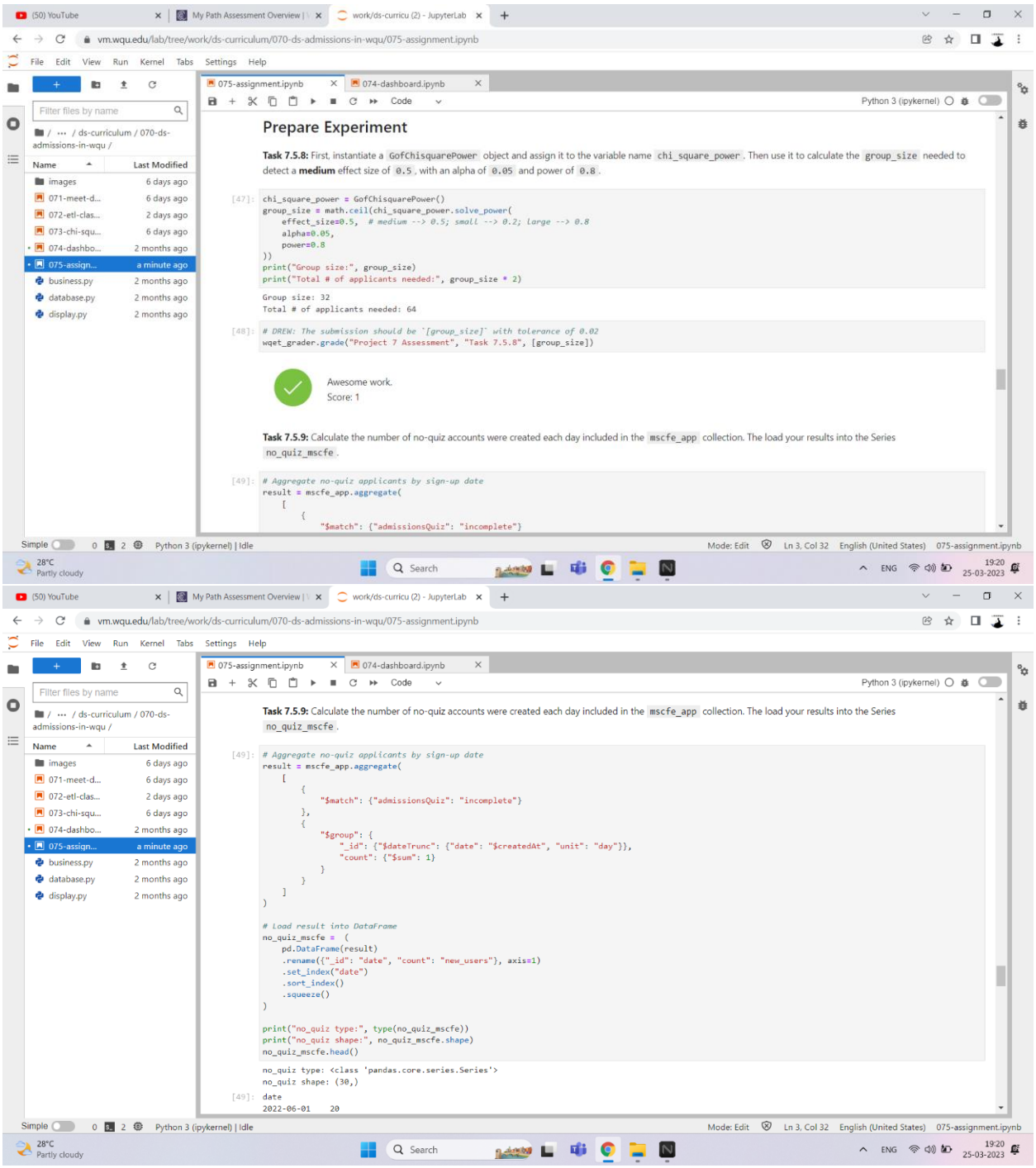

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
def assign_to_groups(self, date_string):
# get observations
observations = self.find_by_date(date_string)
# Shuffle `observations`
random.seed(42)
random.shuffle(observations)
# Get index position of item at observations halfway point
idx = len(observations) // 2
# Assign first half of observations to control group
for doc in observations[:idx]:
doc["inExperiment"] = True
doc["group"] = "no email (control)"
# Assign second half of observations to treatment group
for doc in observations[idx:]:
doc["inExperiment"] = True
doc["group"] = "email (treatment)"
# Update collections
result = self.update_applicants(observations)
return result
# `find_exp_observations` method
def find_exp_observations(self):
result = self.collection.find({"inExperiment": True})
return list(result)
[41]: dir(repo)
[41]: ['__class__',
'__delattr__',
'__dict__',
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:20 25-03-2023


'__dict__',
'__dir__',
'__doc__',
'__eq__',
'__format__',
'__ge__',
'__getattribute__',
'__gt__',
'__hash__',
'__init__',
'__init_subclass__',
'__le__',
'__lt__',
'__module__',
'__ne__',
'__new__',
'__reduce__',
'__reduce_ex__',
'__repr__',
'__setattr__',
'__sizeof__',
'__str__',
'__subclasshook__',
'__weakref__',
'assign_to_groups',
'collection',
'find_by_date',
'find_exp_observations',
'update_applicants']
Task 7.5.5: Create a class definition for your MongoRepository, including an __init__ function that will assign a collection attribute based on user input. Then create an instance of your class named repo. The grader will test whether repo is associated with the correct collection.
[42]:

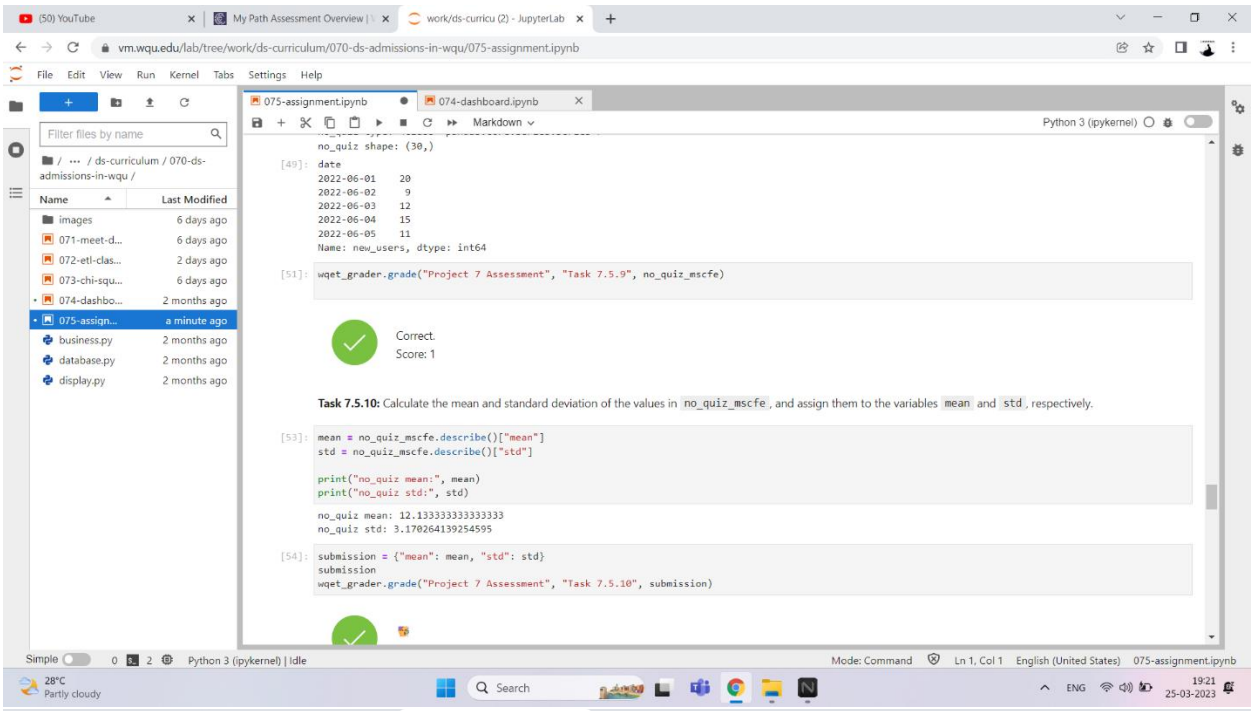

Task 7.5.5: Create a class definition for your MongoRepository, including an __init__ function that will assign a collection attribute based on user input. Then create an instance of your class named repo. The grader will test whether repo is associated with the correct collection.
[42]:
# An instance of class MongoRepository
repo = MongoRepository()
print("repo type:", type(repo))
repo
repo type: <class '__main__.MongoRepository'>
[42]: <__main__.MongoRepository at 0x7f3738a85ac0>
[43]:
submission = {
"is_mongorepo": isinstance(repo, MongoRepository),
"repo_name": repo.collection.name,
}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.5", submission)
Excellent work.
Score: 1
Task 7.5.6: Add a find_by_date method to your class definition for MongoRepository. The method should search the class collection and return all the no-quiz applicants from a specific date. The grader will check your method by looking for applicants whose accounts were created on 1 June 2022.
Warning: Once you update your class definition above, you'll need to rerun that cell and then re-instantiate repo. Otherwise, you won't be able to submit to the grader for this task.

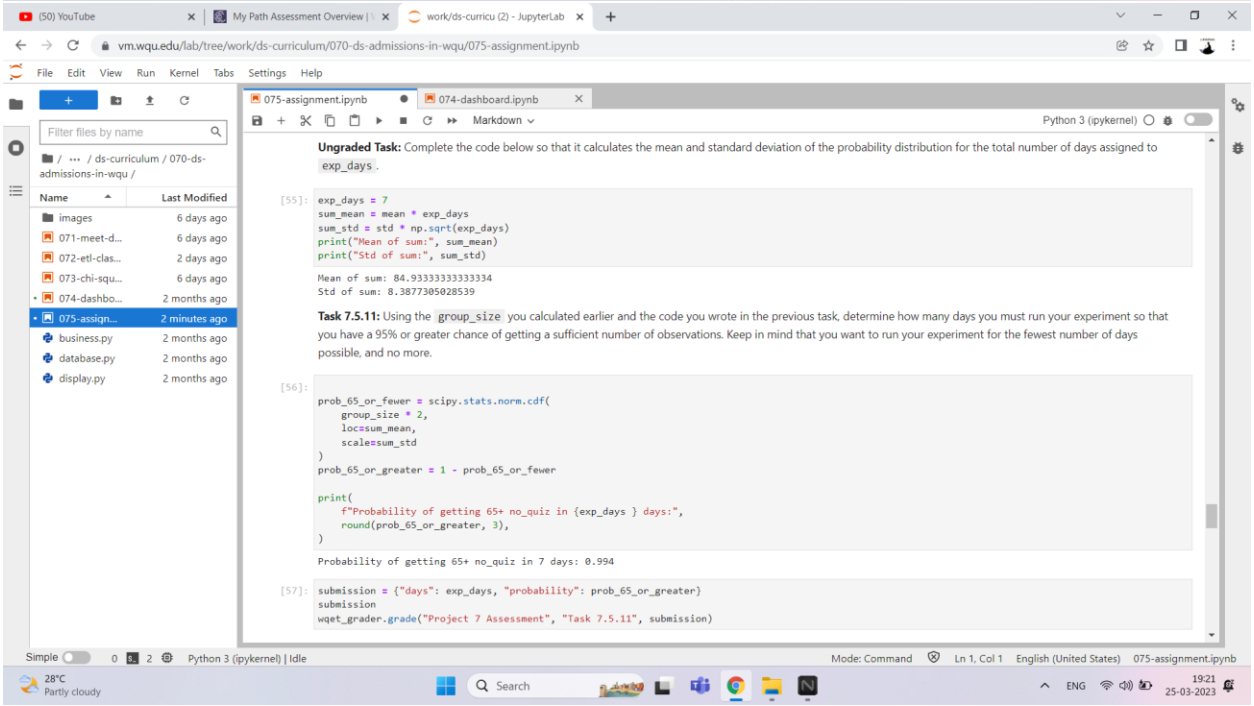

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.6: Add a find_by_date method to your class definition for MongoRepository. The method should search the class collection and return all the no-quiz applicants from a specific date. The grader will check your method by looking for applicants whose accounts were created on 1 June 2022.
Warning: Once you update your class definition above, you'll need to rerun that cell and then re-instantiate repo. Otherwise, you won't be able to submit to the grader for this task.
[45]:
submission = wqet_grader.clean_bson(repo.find_by_date("2022-06-01"))
wqet_grader.grade("Project 7 Assessment", "Task 7.5.6", submission)
Yup. You got it.
Score: 1
Task 7.5.7: Add an assign_to_groups method to your class definition for MongoRepository. It should find users from that date, assign them to groups, update the database, and return the results of the transaction. In order for this method to work, you may also need to create an update_applicants method, too.
Warning: Once you update your class definition above, you'll need to rerun that cell and then re-instantiate repo. Otherwise, you won't be able to submit to the grader for this task.
[46]:
date = "2022-06-02"
repo.assign_to_groups(date)
submission = wqet_grader.clean_bson(repo.find_by_date(date))
wqet_grader.grade("Project 7 Assessment", "Task 7.5.7", submission)
Party time!
Simple
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 3, Col 32
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:20
25-03-2023

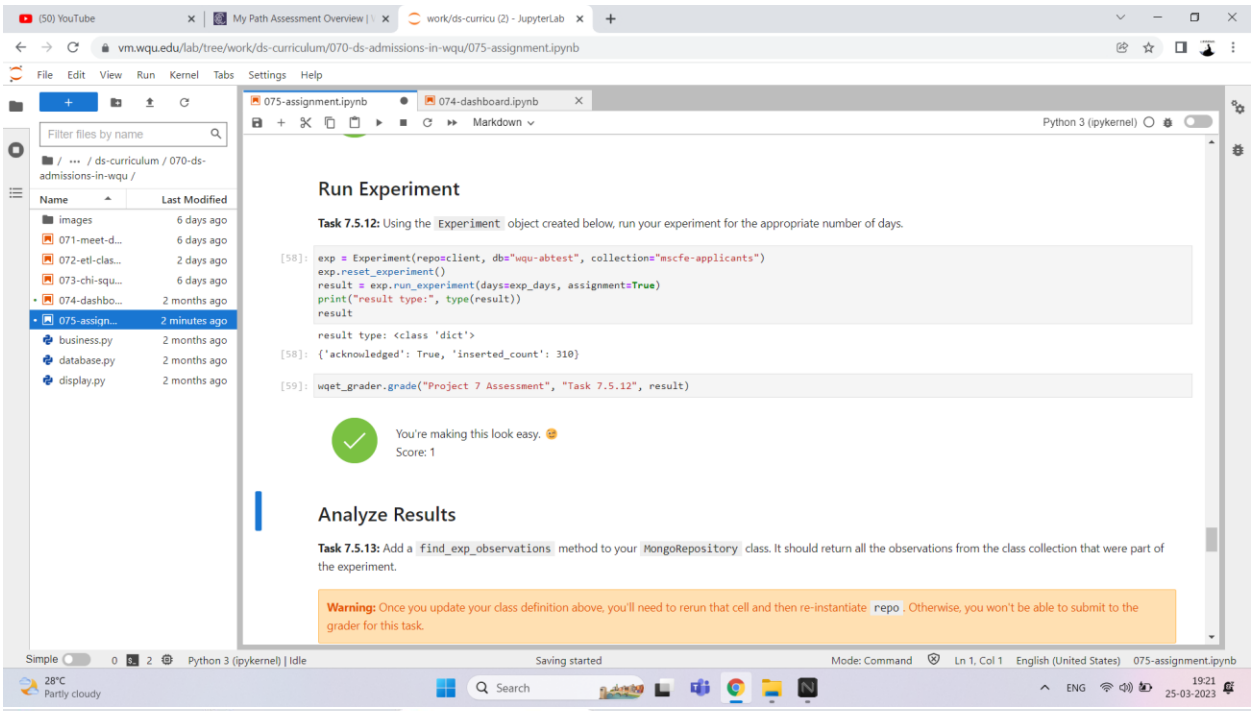

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
[46]:
date = "2022-06-02"
repo.assign_to_groups(date)
submission = wqet_grader.clean_bson(repo.find_by_date(date))
wqet_grader.grade("Project 7 Assessment", "Task 7.5.7", submission)
Party time!
Score: 1
Experiment
Prepare Experiment
Task 7.5.8: First, instantiate a GofChisquarePower object and assign it to the variable name chi_square_power. Then use it to calculate the group_size needed to detect a medium effect size of 0.5, with an alpha of 0.05 and power of 0.8.
[47]:
chi_square_power = GofChisquarePower()
group_size = math.ceil(chi_square_power.solve_power(
effect_size=0.5, # medium --> 0.5; small --> 0.2; large --> 0.8
alpha=0.05,
power=0.8
))
print("Group size:", group_size)
print("Total # of applicants needed:", group_size * 2)
Group size: 32
Total # of applicants needed: 64
[48]:
# DREW: The submission should be `[group_size]` with tolerance of 0.02
Simple
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 3, Col 32
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:20
25-03-2023

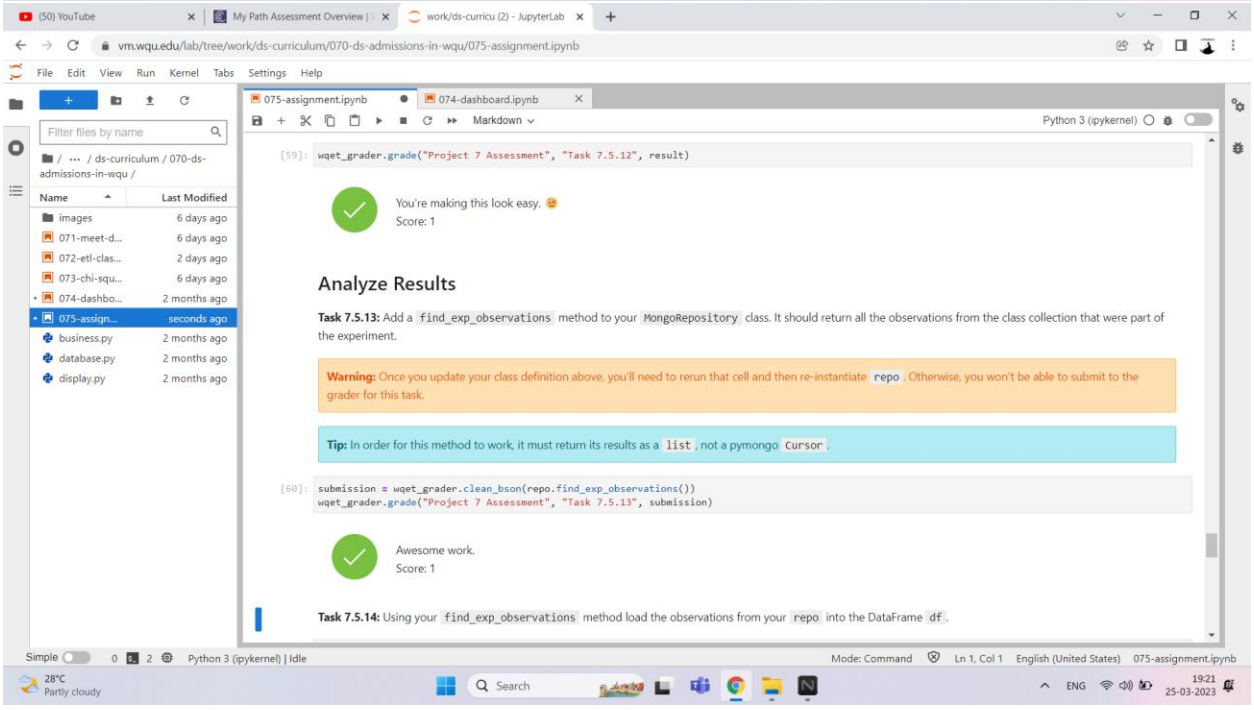

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Prepare Experiment
Task 7.5.8: First, instantiate a GofChisquarePower object and assign it to the variable name chi_square_power. Then use it to calculate the group_size needed to detect a medium effect size of 0.5, with an alpha of 0.05 and power of 0.8.
[47]:
chi_square_power = GofChisquarePower()
group_size = math.ceil(chi_square_power.solve_power(
effect_size=0.5, # medium --> 0.5; small --> 0.2; large --> 0.8
alpha=0.05,
power=0.8
))
print("Group size:", group_size)
print("Total # of applicants needed:", group_size * 2)
Group size: 32
Total # of applicants needed: 64
[48]:
# DREW: The submission should be `[group_size]` with tolerance of 0.02
wqet_grader.grade("Project 7 Assessment", "Task 7.5.8", [group_size])
Awesome work.
Score: 1
Task 7.5.9: Calculate the number of no-quiz accounts were created each day included in the mscfe_app collection. The load your results into the Series no_quiz_mscfe.
[49]:
# Aggregate no-quiz applicants by sign-up date
result = mscfe_app.aggregate(
[
{
"$match": {"admissionsQuiz": "incomplete"}
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Edit Ln 3, Col 32 English (United States) 075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:20
25-03-2023

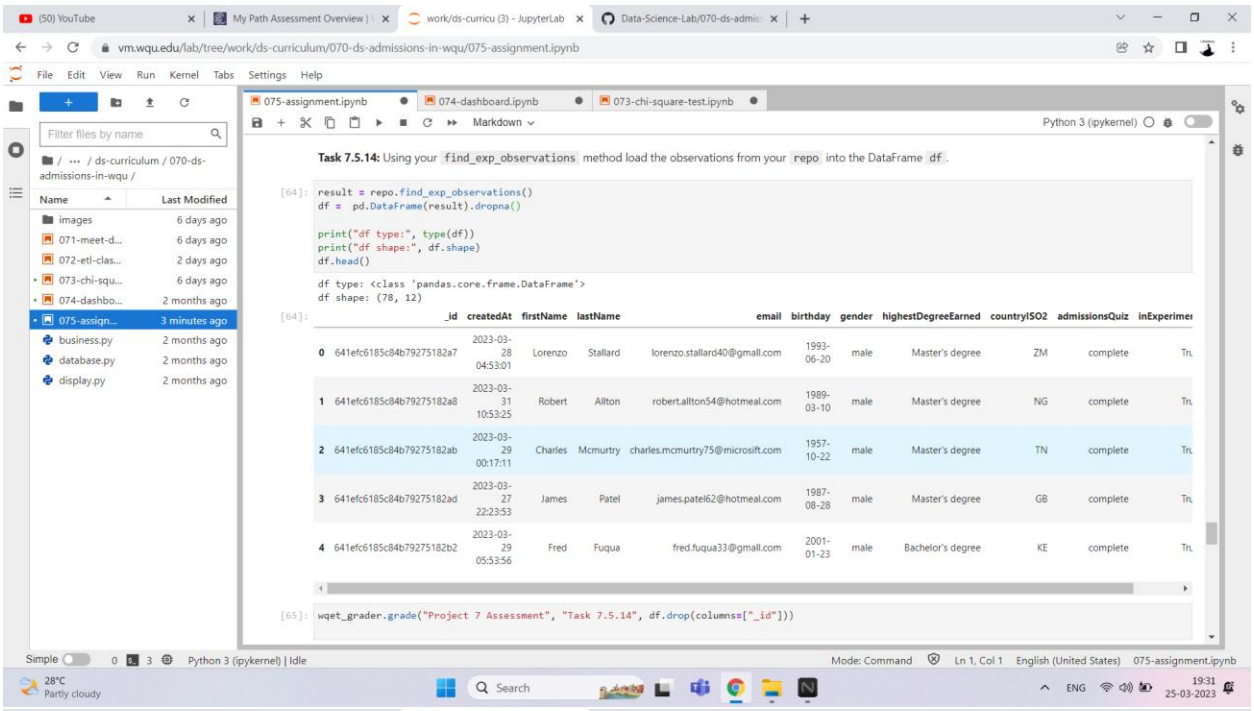

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Code
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.9: Calculate the number of no-quiz accounts were created each day included in the mscfe_app collection. The load your results into the Series no_quiz_mscfe.
[49]:
# Aggregate no-quiz applicants by sign-up date
result = mscfe_app.aggregate(
[
{
"$match": {"admissionsQuiz": "incomplete"}
},
{
"$group": {
"_id": {"$dateTrunc": {"date": "$createdAt", "unit": "day"}},
"count": {"$sum": 1}
}
}
]
)
# Load result into DataFrame
no_quiz_mscfe = (
pd.DataFrame(result)
.rename({"_id": "date", "count": "new_users"}, axis=1)
.set_index("date")
.sort_index()
.squeeze()
)
print("no_quiz type:", type(no_quiz_mscfe))
print("no_quiz shape:", no_quiz_mscfe.shape)
no_quiz_mscfe.head()
no_quiz type: <class 'pandas.core.series.Series'>
no_quiz shape: (30,)
[49]:
date
2022-06-01 20
Simple 0 2 Python 3 (ipykernel) | Idle
Mode: Edit Ln 3, Col 32 English (United States) 075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:20
25-03-2023

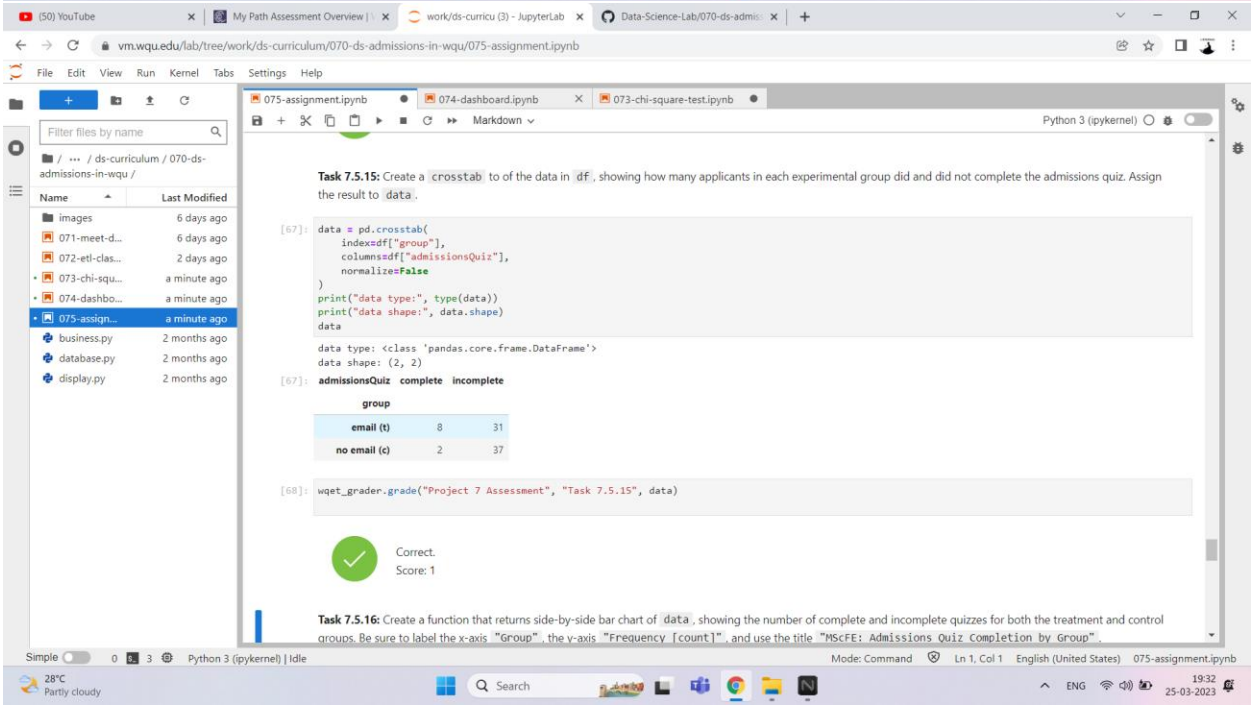

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
no_quiz shape: (30,)
[49]:
date
2022-06-01 20
2022-06-02 9
2022-06-03 12
2022-06-04 15
2022-06-05 11
Name: new_users, dtype: int64
[51]:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.9", no_quiz_mscfe)
Correct.
Score: 1
Task 7.5.10: Calculate the mean and standard deviation of the values in no_quiz_mscfe, and assign them to the variables mean and std, respectively.
[53]:
mean = no_quiz_mscfe.describe()["mean"]
std = no_quiz_mscfe.describe()["std"]
print("no_quiz mean:", mean)
print("no_quiz std:", std)
no_quiz mean: 12.133333333333333
no_quiz std: 3.170264139254595
[54]:
submission = {"mean": mean, "std": std}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.10", submission)
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:21
25-03-2023

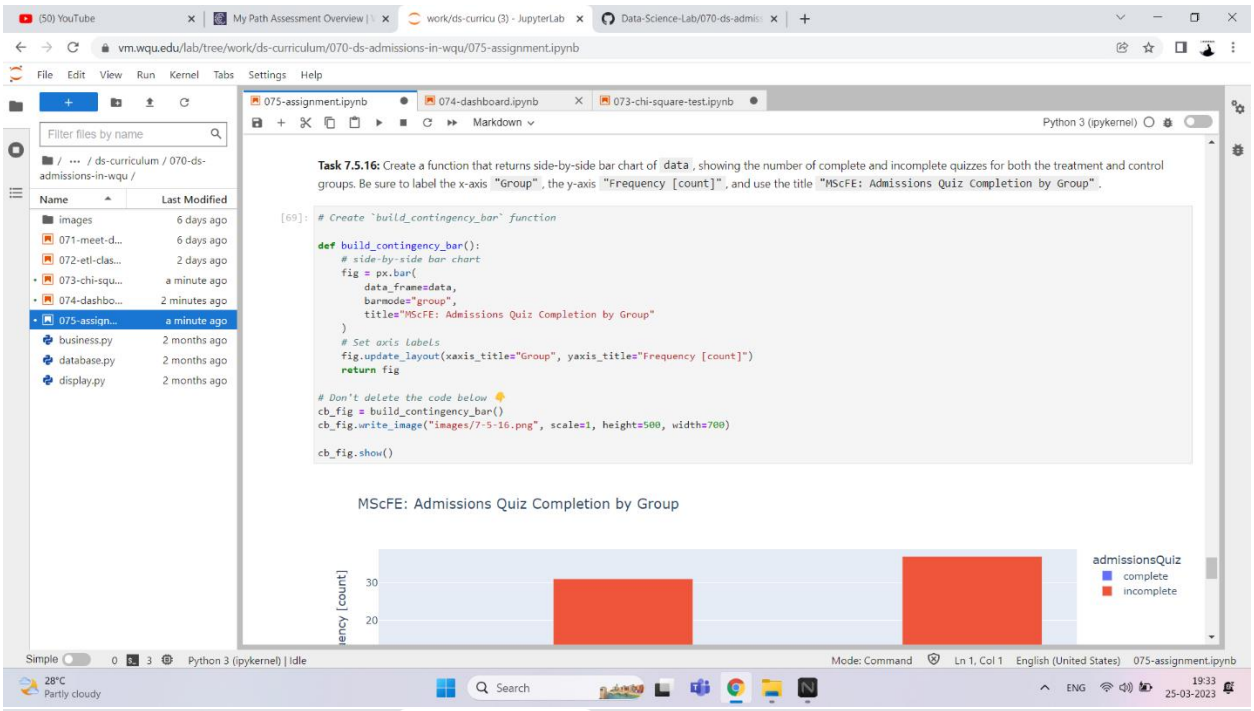

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... 2 minutes ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Ungraded Task: Complete the code below so that it calculates the mean and standard deviation of the probability distribution for the total number of days assigned to exp_days.
[55]:
exp_days = 7
sum_mean = mean * exp_days
sum_std = std * np.sqrt(exp_days)
print("Mean of sum:", sum_mean)
print("Std of sum:", sum_std)
Mean of sum: 84.93333333333334
Std of sum: 8.3877305028539
Task 7.5.11: Using the group_size you calculated earlier and the code you wrote in the previous task, determine how many days you must run your experiment so that you have a 95% or greater chance of getting a sufficient number of observations. Keep in mind that you want to run your experiment for the fewest number of days possible, and no more.
[56]:
prob_65_or_fewer = scipy.stats.norm.cdf(
group_size * 2,
loc=sum_mean,
scale=sum_std
)
prob_65_or_greater = 1 - prob_65_or_fewer
print(
f"Probability of getting 65+ no_quiz in {exp_days } days:",
round(prob_65_or_greater, 3),
)
Probability of getting 65+ no_quiz in 7 days: 0.994
[57]:
submission = {"days": exp_days, "probability": prob_65_or_greater}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.11", submission)
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:21
25-03-2023

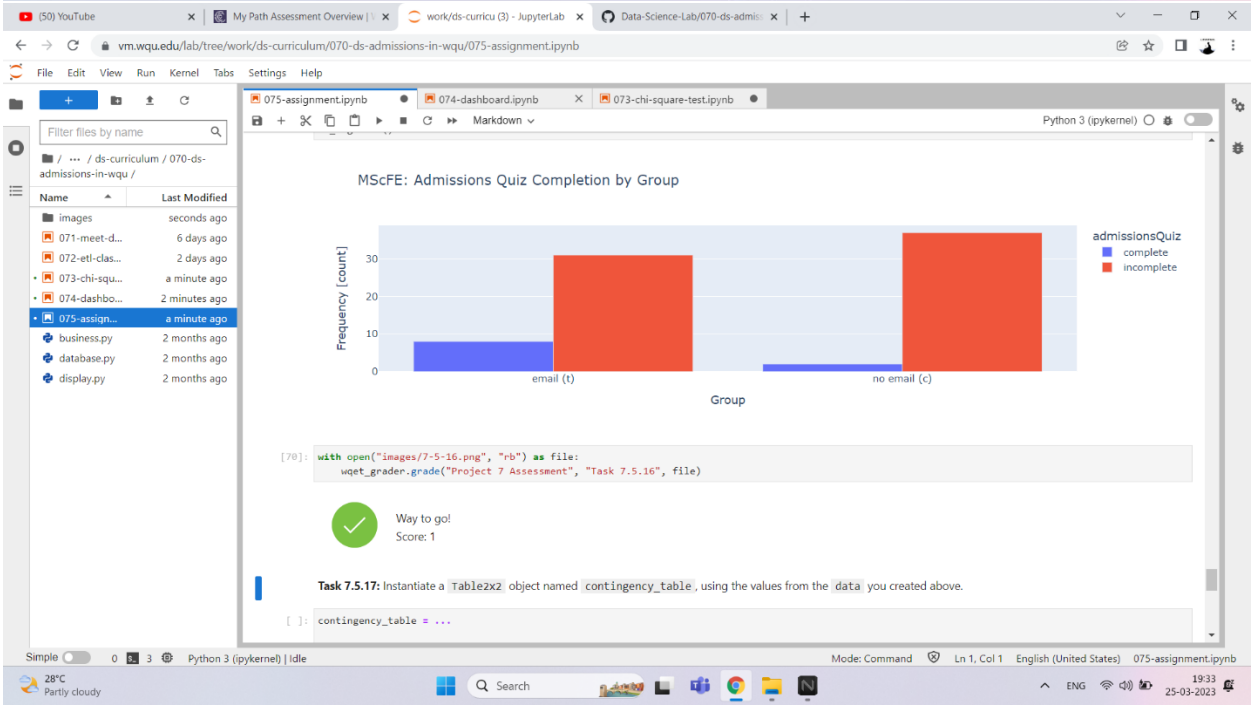

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... 2 minutes ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Run Experiment
Task 7.5.12: Using the Experiment object created below, run your experiment for the appropriate number of days.
[58]:
exp = Experiment(repo=client, db="wqu-abtest", collection="mscfe-applicants")
exp.reset_experiment()
result = exp.run_experiment(days=exp_days, assignment=True)
print("result type:", type(result))
result
result type: <class 'dict'>
[58]:
{'acknowledged': True, 'inserted_count': 310}
[59]:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.12", result)
You're making this look easy.
Score: 1
Analyze Results
Task 7.5.13: Add a find_exp_observations method to your MongoRepository class. It should return all the observations from the class collection that were part of the experiment.
Warning: Once you update your class definition above, you'll need to rerun that cell and then re-instantiate repo. Otherwise, you won't be able to submit to the grader for this task.
Simple
Python 3 (ipykernel) | Idle
Saving started
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:21
25-03-2023

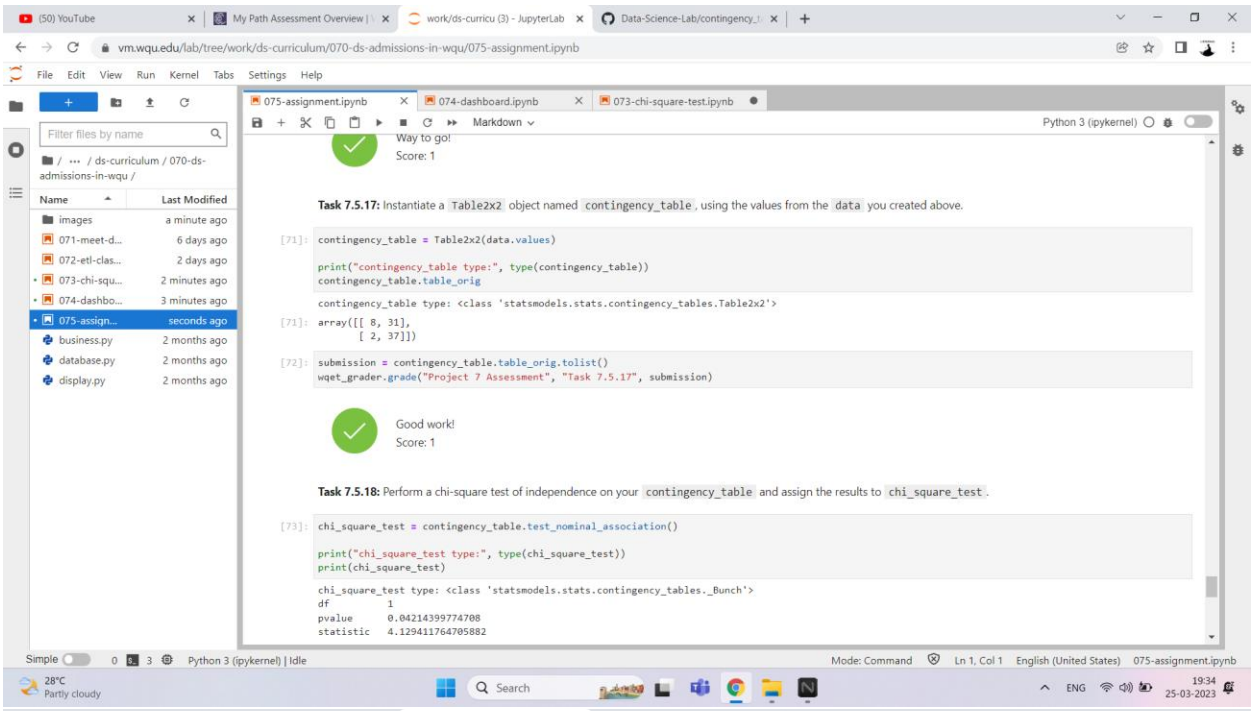

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (2) - JupyterLab
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
[59]:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.12", result)
You're making this look easy.
Score: 1
Analyze Results
Task 7.5.13: Add a find_exp_observations method to your MongoRepository class. It should return all the observations from the class collection that were part of the experiment.
Warning: Once you update your class definition above, you'll need to rerun that cell and then re-instantiate repo. Otherwise, you won't be able to submit to the grader for this task.
Tip: In order for this method to work, it must return its results as a list, not a pymongo Cursor.
[60]:
submission = wqet_grader.clean_bson(repo.find_exp_observations())
wqet_grader.grade("Project 7 Assessment", "Task 7.5.13", submission)
Awesome work.
Score: 1
Task 7.5.14: Using your find_exp_observations method load the observations from your repo into the DataFrame df.
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:21
25-03-2023

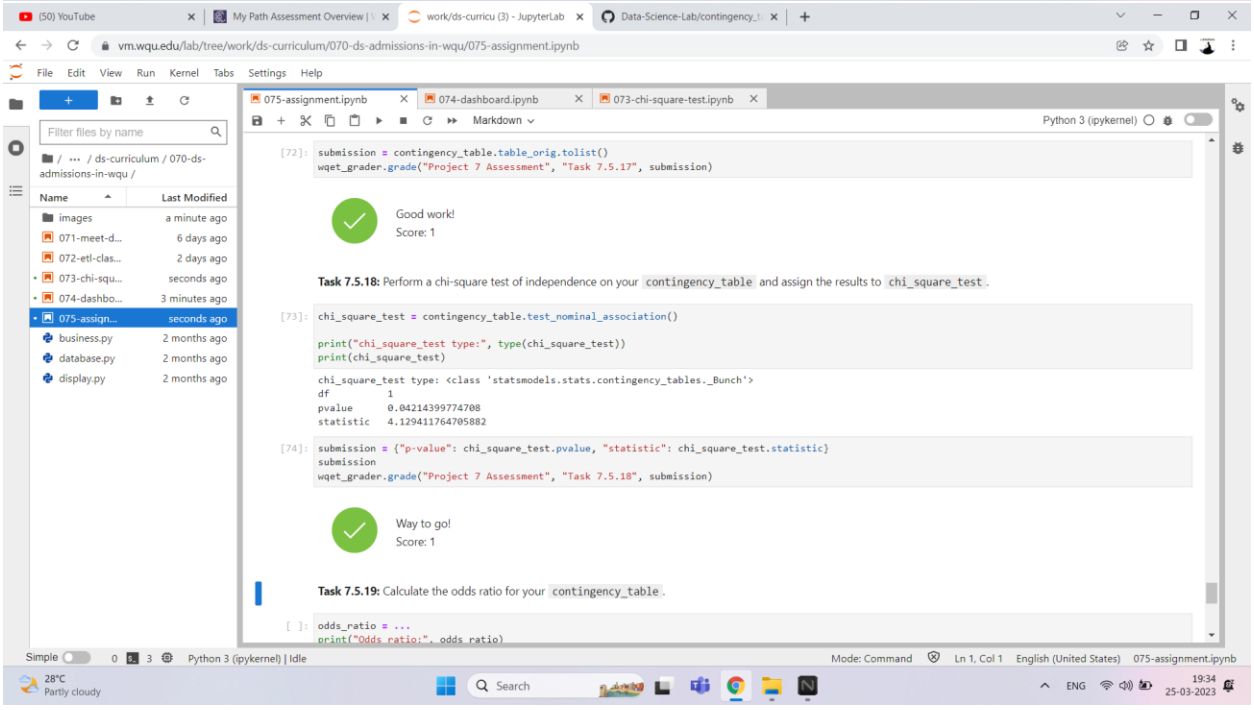

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/070-ds-admiss
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 6 days ago
074-dashbo... 2 months ago
075-assign... 3 minutes ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.14: Using your find_exp_observations method load the observations from your repo into the DataFrame df.
[64]:
result = repo.find_exp_observations()
df = pd.DataFrame(result).dropna()
print("df type:", type(df))
print("df shape:", df.shape)
df.head()
df type: <class 'pandas.core.frame.DataFrame'>
df shape: (78, 12)
[64]:
_id createdAt firstName lastName email birthday gender highestDegreeEarned countryISO2 admissionsQuiz inExperimer
0 641efc6185c84b79275182a7 2023-03-28 04:53:01 Lorenzo Stallard lorenzo.stallard40@gmail.com 1993-06-20 male Master's degree ZM complete Tru
1 641efc6185c84b79275182a8 2023-03-31 10:53:25 Robert Allton robert.allton54@hotmeal.com 1989-03-10 male Master's degree NG complete Tru
2 641efc6185c84b79275182ab 2023-03-29 00:17:11 Charles Mcmurtry charles.mcmurtry75@microsift.com 1957-10-22 male Master's degree TN complete Tru
3 641efc6185c84b79275182ad 2023-03-27 22:23:53 James Patel james.patel62@hotmeal.com 1987-08-28 male Master's degree GB complete Tru
4 641efc6185c84b79275182b2 2023-03-29 05:53:56 Fred Fuqua fred.fuqua33@gmail.com 2001-01-23 male Bachelor's degree KE complete Tru
[65]:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.14", df.drop(columns=["_id"]))
Simple 0 3 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:31 25-03-2023

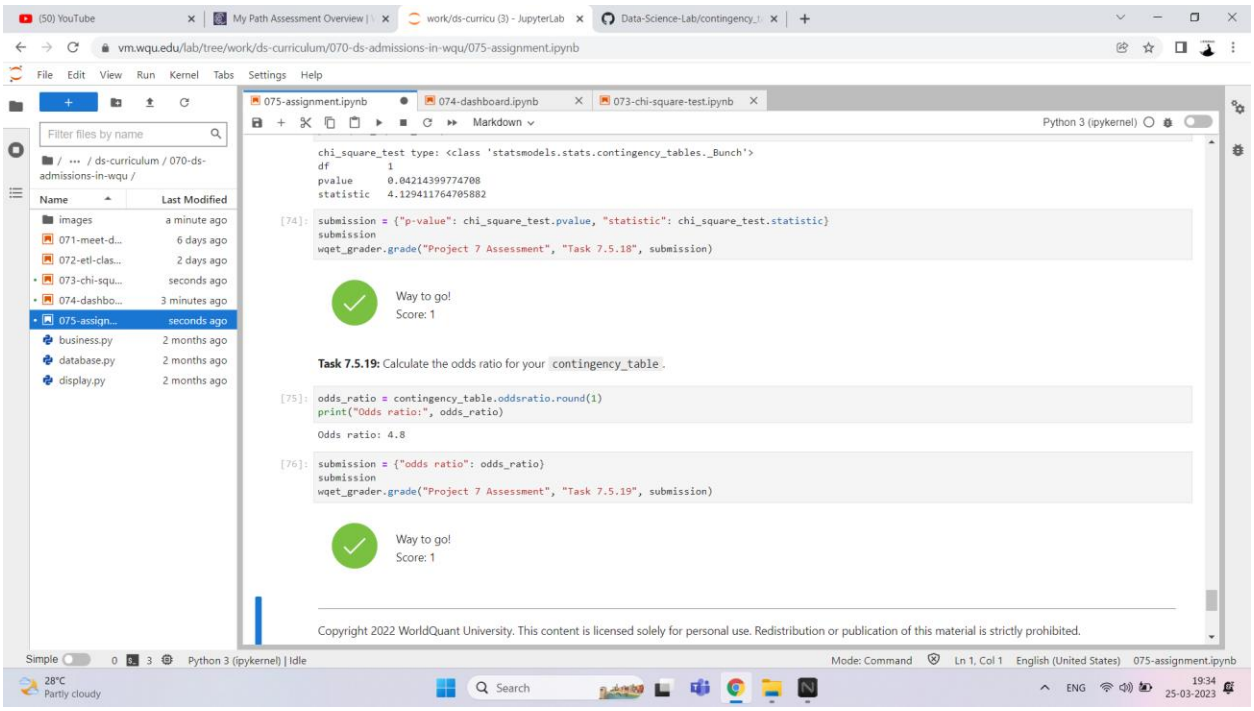

(50) YouTube
My Path Assessment Overview |
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/070-ds-admiss
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... a minute ago
074-dashbo... a minute ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.15: Create a crosstab to of the data in df, showing how many applicants in each experimental group did and did not complete the admissions quiz. Assign the result to data.
[67]:
data = pd.crosstab(
index=df["group"],
columns=df["admissionsQuiz"],
normalize=False
)
print("data type:", type(data))
print("data shape:", data.shape)
data
data type: <class 'pandas.core.frame.DataFrame'>
data shape: (2, 2)
[67]:
admissionsQuiz complete incomplete
group
email (t) 8 31
no email (c) 2 37
[68]:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.15", data)
Correct.
Score: 1
Task 7.5.16: Create a function that returns side-by-side bar chart of data, showing the number of complete and incomplete quizzes for both the treatment and control groups. Be sure to label the x-axis "Group", the y-axis "Frequency [count]", and use the title "MScFE: Admissions Quiz Completion by Group".
Simple 0 3 Python 3 (ipykernel) | Idle
Mode: Command Ln 1, Col 1 English (United States) 075-assignment.ipynb
28°C Partly cloudy
Search
ENG 19:32 25-03-2023


(50) YouTube
My Path Assessment Overview
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/070-ds-admiss
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images 6 days ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... a minute ago
074-dashbo... 2 minutes ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Task 7.5.16: Create a function that returns side-by-side bar chart of data, showing the number of complete and incomplete quizzes for both the treatment and control groups. Be sure to label the x-axis "Group", the y-axis "Frequency [count]", and use the title "MScFE: Admissions Quiz Completion by Group".
[69]:
# Create `build_contingency_bar` function
def build_contingency_bar():
# side-by-side bar chart
fig = px.bar(
data_frame=data,
barmode="group",
title="MScFE: Admissions Quiz Completion by Group"
)
# Set axis labels
fig.update_layout(xaxis_title="Group", yaxis_title="Frequency [count]")
return fig
# Don't delete the code below
cb_fig = build_contingency_bar()
cb_fig.write_image("images/7-5-16.png", scale=1, height=500, width=700)
cb_fig.show()
MScFE: Admissions Quiz Completion by Group
admissionsQuiz
complete
incomplete
30
20
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:33
25-03-2023


(50) YouTube
My Path Assessment Overview
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/070-ds-admiss
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images seconds ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... a minute ago
074-dashbo... 2 minutes ago
075-assign... a minute ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
MScFE: Admissions Quiz Completion by Group
admissionsQuiz
complete
incomplete
Frequency [count]
30
20
10
0
email (t)
no email (c)
Group
[70]:
with open("images/7-5-16.png", "rb") as file:
wqet_grader.grade("Project 7 Assessment", "Task 7.5.16", file)
Way to go!
Score: 1
Task 7.5.17: Instantiate a Table2x2 object named contingency_table, using the values from the data you created above.
[ ]:
contingency_table = ...
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:33
25-03-2023


(50) YouTube
My Path Assessment Overview |
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/contingency_t
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images a minute ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... 2 minutes ago
074-dashbo... 3 minutes ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
Way to go!
Score: 1
Task 7.5.17: Instantiate a Table2x2 object named contingency_table, using the values from the data you created above.
[71]: contingency_table = Table2x2(data.values)
print("contingency_table type:", type(contingency_table))
contingency_table.table_orig
contingency_table type: <class 'statsmodels.stats.contingency_tables.Table2x2'>
[71]: array([[ 8, 31],
[ 2, 37]])
[72]: submission = contingency_table.table_orig.tolist()
wqet_grader.grade("Project 7 Assessment", "Task 7.5.17", submission)
Good work!
Score: 1
Task 7.5.18: Perform a chi-square test of independence on your contingency_table and assign the results to chi_square_test.
[73]: chi_square_test = contingency_table.test_nominal_association()
print("chi_square_test type:", type(chi_square_test))
print(chi_square_test)
chi_square_test type: <class 'statsmodels.stats.contingency_tables._Bunch'>
df 1
pvalue 0.04214399774708
statistic 4.129411764705882
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:34
25-03-2023


(50) YouTube
My Path Assessment Overview |
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/contingency_t
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
Filter files by name
/ ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images a minute ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... seconds ago
074-dashbo... 3 minutes ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
[72]: submission = contingency_table.table_orig.tolist()
wqet_grader.grade("Project 7 Assessment", "Task 7.5.17", submission)
Good work!
Score: 1
Task 7.5.18: Perform a chi-square test of independence on your contingency_table and assign the results to chi_square_test.
[73]: chi_square_test = contingency_table.test_nominal_association()
print("chi_square_test type:", type(chi_square_test))
print(chi_square_test)
chi_square_test type: <class 'statsmodels.stats.contingency_tables._Bunch'>
df 1
pvalue 0.04214399774708
statistic 4.129411764705882
[74]: submission = {"p-value": chi_square_test.pvalue, "statistic": chi_square_test.statistic}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.18", submission)
Way to go!
Score: 1
Task 7.5.19: Calculate the odds ratio for your contingency_table.
[ ]: odds_ratio = ...
print("Odds ratio:", odds_ratio)
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:34
25-03-2023


(50) YouTube
My Path Assessment Overview |
work/ds-curricu (3) - JupyterLab
Data-Science-Lab/contingency_t
vm.wqu.edu/lab/tree/work/ds-curriculum/070-ds-admissions-in-wqu/075-assignment.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ ... / ds-curriculum / 070-ds-admissions-in-wqu /
Name
Last Modified
images a minute ago
071-meet-d... 6 days ago
072-etl-clas... 2 days ago
073-chi-squ... seconds ago
074-dashbo... 3 minutes ago
075-assign... seconds ago
business.py 2 months ago
database.py 2 months ago
display.py 2 months ago
075-assignment.ipynb
074-dashboard.ipynb
073-chi-square-test.ipynb
Markdown
Python 3 (ipykernel)
chi_square_test type: <class 'statsmodels.stats.contingency_tables._Bunch'>
df 1
pvalue 0.04214399774708
statistic 4.129411764705882
[74]:
submission = {"p-value": chi_square_test.pvalue, "statistic": chi_square_test.statistic}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.18", submission)
Way to go!
Score: 1
Task 7.5.19: Calculate the odds ratio for your contingency_table.
[75]:
odds_ratio = contingency_table.oddsratio.round(1)
print("Odds ratio:", odds_ratio)
Odds ratio: 4.8
[76]:
submission = {"odds ratio": odds_ratio}
submission
wqet_grader.grade("Project 7 Assessment", "Task 7.5.19", submission)
Way to go!
Score: 1
Copyright 2022 WorldQuant University. This content is licensed solely for personal use. Redistribution or publication of this material is strictly prohibited.
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
English (United States)
075-assignment.ipynb
28°C
Partly cloudy
Search
ENG
19:34
25-03-2023