Data Science Immersive

General Assembly

Links in Slack, Spring 2024

Matthew Morris

TAs: John Hazard, Kihoon Sohn

A diagram of a flowchart

Description automatically generated

Lectures

[01 Orientation and Review home, slides (12/12) 3](#_Toc158659088)

[Other Gen’l. Python intro 4](#_Toc158659089)

[02 Development Environment home, slides (12/14) 7](#_Toc158659090)

[03 Jupyter Numpy Pandas home, slides (12/19) 10](#_Toc158659091)

[04 Lab Presentations Catchup slides (12/21) 11](#_Toc158659092)

[05 Intro Exploratory Data Analysis (EDA) in Pandas (1/2/24) 12](#_Toc158659093)

[06 Statistics in Python (R 1/4) 14](#_Toc158659094)

[07 More EDA Data Visualization in Python (T 1/9) 18](#_Toc158659095)

[08 Experiments and Hypothesis Testing (R 1/11) 19](#_Toc158659096)

[09 Presentations (T 1/16) 21](#_Toc158659097)

[10 KNN/ Classification (R 1/18) 22](#_Toc158659098)

[11 Train-Test Split & Bias Variance (T 1/23) 23](#_Toc158659099)

[12 Linear Regression (R 1/25) 24](#_Toc158659100)

[13 Logistic Regression (1/30) 25](#_Toc158659101)

[14 Presentations (R 2/1) 26](#_Toc158659102)

[15 Presentations, Time Series (T 2/6) 27](#_Toc158659103)

[16 Time Series, APIs (R 2/8) 28](#_Toc158659104)

[17 NLP (T 2/13) 29](#_Toc158659105)

[18 Flex subject and Class time 30](#_Toc158659106)

[19 Flex day, review, catchup, workshop 31](#_Toc158659107)

# 01 Orientation and Review home, slides (12/12)

<https://www.data-is-plural.com/>

A black text on a white background

Description automatically generated

<https://www.kaggle.com/datasets>

A close-up of a text

Description automatically generated

<https://realpython.com/using-jupyterlab/>

A cartoon of a person in a lab coat

Description automatically generated

<https://dataschool.ck.page/posts/tuesday-tip-34-what-are-conda-anaconda-and-miniconda>

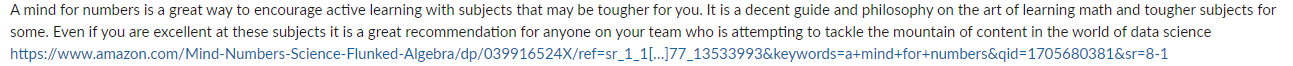
A reptile with text

Description automatically generated

## Other Gen’l. Python intro

From Jan 19th class, but Python Generals

A Mind for Numbers [Amazon](https://www.amazon.com/Mind-Numbers-Science-Flunked-Algebra/dp/039916524X/ref=sr_1_1?dib=eyJ2IjoiMSJ9.SPkmzBjZVMsan3na5qxs6sqNkxGzdt1mifiIfzKsO9yaJZM1IvtxE_B3EPkgmgINmceyM3GwJrgXbkxDFHRBXpFMWY5tCzIe7V6RpN0Yxni4cb2bIwLIvlHI5z4KJanaDMdB0R-0aqJiScxbLj1IxNcliD6BU0vqkw41teg7fA9XCV-L1WbqEPpx__xY_hl0Fmb87MlIL2-i5l1AXk7jMCPh239DJBhJZRdMi1-x_NU.tT5XYH1IoLvY2WXj4H4ld6R51lAKG986JyrXXNhJrus&dib_tag=se&hvadid=580626274215&hvdev=c&hvlocphy=9033619&hvnetw=g&hvqmt=e&hvrand=11671802313835774148&hvtargid=kwd-69728463346&hydadcr=3177_13533993&keywords=a+mind+for+numbers&qid=1705680381&sr=8-1)



Numsense: No Math Added [Amazon](https://www.goodreads.com/en/book/show/34213247)

A screenshot of a computer

Description automatically generated

[YouTube](https://www.youtube.com/watch?v=F9kC70vhRnw): Learning Python the Hard Way and [AutomateTheBoringStuff](https://automatetheboringstuff.com/)

A close-up of a computer screen

Description automatically generated

[Goodreads](https://www.goodreads.com/book/show/60704827-dive-into-data-science): Dive into Data Science

A screenshot of a computer

Description automatically generated

DMBOK Data Management book [Amazon](https://www.amazon.com/DAMA-DMBOK-Data-Management-Body-Knowledge/dp/1634622340/ref=sr_1_1?dib=eyJ2IjoiMSJ9.i3wVi6cE994TwH73g40MsCDzhYAGAJhCIqKFy_Zem1_tVlZvKWXF8MWvio2u8lLbGAngCzT26Wtch-XaoPNWm0rSV-voalW6kFT9lVAjXnRJOMcIkebw9AbanxB8Rw4HVioOSMwQ2wwQUVM4-Xim1ENZrt27iV8JLzBfrMW8REeJWZahBi2wGugpce3hA_ZY78nll9_JwFrZTb1nn22EeRYMSoGHON5i2iZt-b4N2cw.kfLK-Ku_3fgq2TJuGDX1gmIZBfuJZ3pSrjf8pT47PxM&dib_tag=se&hvadid=598612844310&hvdev=c&hvlocphy=9033619&hvnetw=g&hvqmt=e&hvrand=9330558952115290709&hvtargid=kwd-13224844317&hydadcr=9655_13572754&keywords=dmbok&qid=1705681207&sr=8-1&ufe=app_do%3Aamzn1.fos.006c50ae-5d4c-4777-9bc0-4513d670b6bc)

Data Governance [GoodReads](https://www.goodreads.com/book/show/49086278-data-governance?from_search=true&from_srp=true&qid=G0HUJNUOSS&rank=2)

A screenshot of a computer

Description automatically generated

[Datawarehousing](https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-lifecycle-toolkit/)

O’Reilly SQL [Hacks](https://www.oreilly.com/library/view/sql-hacks/0596527993/)

A screenshot of a computer

Description automatically generated

# 02 Development Environment home, slides (12/14)

<https://en.wikipedia.org/wiki/Discretization#:~:text=In%20applied%20mathematics%2C%20discretization%20is,and%20equations%20into%20discrete%20counterparts>

A screenshot of a computer

Description automatically generated

Discrete = Dimensions, Continuous = Measures:

A screenshot of a computer

Description automatically generated

<https://realpython.com/lessons/scripts-modules-packages-and-libraries/#:~:text=A%20script%20is%20a%20Python,into%20scripts%20or%20other%20modules>

A cartoon of a person working on a computer

Description automatically generated

<https://docs.spyder-ide.org/current/videos/first-steps-with-spyder.html>

A white background with black text

Description automatically generated

# 03 Jupyter Numpy Pandas home, slides (12/19)

YouTube – Matthew Python [fundamentals](https://www.youtube.com/watch?v=4_7V3dzEWF0&list=PL-DpXV0x2lV_Ocr5DLN9R-f60j-4rXLou&ab_channel=SharedXP)

A close-up of a computer screen

Description automatically generated

# 04 Lab Presentations Catchup slides (12/21)

# 05 Intro Exploratory Data Analysis (EDA) in Pandas (1/2/24)

<https://www.youtube.com/watch?v=5Zg-C8AAIGg>

A white background with black text

Description automatically generated

<https://www.youtube.com/watch?v=hVimVzgtD6w>

A white calendar with a person holding a pen

Description automatically generated with medium confidence

[First Step toward Missing Data Imputation must never be Imputation](https://www.blog.dailydoseofds.com/p/the-first-step-towards-missing-data?utm_source=post-email-title&publication_id=1119889&post_id=139932392&utm_campaign=email-post-title&isFreemail=true&r=w9uj&utm_medium=email)

A screenshot of a computer

Description automatically generated

A diagram of a normal distribution

Description automatically generated

# 06 Statistics in Python (R 1/4)

<https://machinelearning.cards/>

A screenshot of a computer

Description automatically generated

Missing Value [Imputation](https://scikit-learn.org/stable/modules/impute.html)

A screenshot of a computer

Description automatically generated

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rank.html>

A close-up of a list of text

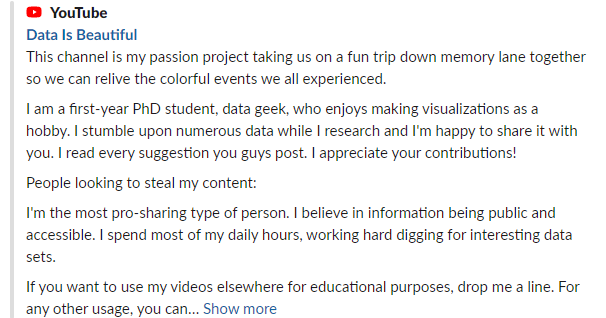
Description automatically generated

[YouTube](https://www.youtube.com/watch?v=QFrqTFRy-LU)

A screenshot of a computer

Description automatically generated

Data is [Beautiful](https://www.youtube.com/@DataIsBeautifulOfficial/videos)



AWS Certified Cloud Practitioner Fundamentals [exam](https://aws.amazon.com/certification/certified-cloud-practitioner/) (relatively easy)

A close-up of a certificate

Description automatically generated

<https://www.statisticshowto.com/choose-bin-sizes-statistics/>

A group of blue trash cans

Description automatically generated

# 07 More EDA Data Visualization in Python (T 1/9)

A screenshot of a computer screen

Description automatically generated

*Tanay can posted the .zip with screenshots of the ipynb along with source code*

# 08 Experiments and Hypothesis Testing (R 1/11)

[Scale article](https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02)

A green lizard with black text

Description automatically generated

Scipy [stats](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html) One-way Anova test

Penguins dataset from [Kaggle](https://www.kaggle.com/datasets/parulpandey/palmer-archipelago-antarctica-penguin-data)

A screenshot of a penguin diagram

Description automatically generated

[YouTube](https://www.youtube.com/watch?v=RBzJR4Emxvo) - Wine Tasting

A person with a beard

Description automatically generated with medium confidence

<http://listen.hatnote.com/>

A close up of a computer screen

Description automatically generated

# 09 Presentations (T 1/16)

Python Recommender [Systems](https://www.datacamp.com/tutorial/recommender-systems-python)

A screenshot of a computer

Description automatically generated

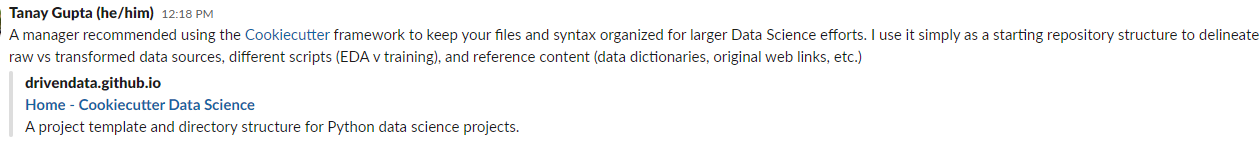
# 10 KNN/ Classification (R 1/18)

Increase Python cell [width](https://m.youtube.com/watch?v=hR9YtVAZSi8)

A screenshot of a computer

Description automatically generated

Cookie cutter file [structure](https://drivendata.github.io/cookiecutter-data-science/#example)



# 11 Train-Test Split & Bias Variance (T 1/23)

# 12 Linear Regression (R 1/25)

Multi [collinearity](https://www.statology.org/multicollinearity-regression/)

A screenshot of a computer

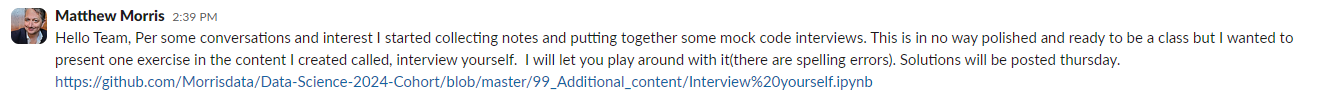
Description automatically generated

# 13 Logistic Regression (1/30)

# 14 Presentations (R 2/1)

# 15 Presentations, Time Series (T 2/6)

Mock Code interview [ipynb](https://github.com/Morrisdata/Data-Science-2024-Cohort/blob/master/99_Additional_content/Interview%20yourself.ipynb)



# 16 Time Series, APIs (R 2/8)

P, D, Q for [ARIMA](https://analyticsindiamag.com/quick-way-to-find-p-d-and-q-values-for-arima/)

A screenshot of a computer

Description automatically generated

Identifying the numbers of AR, MA ([Duke](https://people.duke.edu/~rnau/411arim3.htm))

# 17 NLP (T 2/13)

# 18 Flex subject and Class time

# 19 Flex day, review, catchup, workshop