**Project - Programming for Data Analysis**

Potential Datasets

Employment and Disability

[1] Central Statistics Office, People with a disability, (2021) <https://data.cso.ie/table/APD05>

[2] Towards Data Science, How to read CSV File into Python using Pandas, (25th May, 2020) <https://towardsdatascience.com/how-to-read-csv-file-using-pandas-ab1f5e7e7b58?gi=df516eec2150>

[3] W3 Schools, Pandas Read CSV, (1999 - 2021) <https://www.w3schools.com/python/pandas/pandas_csv.asp>

[4] Data Independent, Pandas Head – Preview Data – DataFrame.head(), (13th Sept, 2020) <https://www.dataindependent.com/pandas/pandas-head/>

[5] Towards Data Science, Pandas for People In A Hurry (23th May, 2019) <https://towardsdatascience.com/pandas-for-people-in-a-hurry-59d966630ae0#:~:text=df.,Example%3A%20df>.

[6] Additional CSO Data Review:

- <https://www.cso.ie/en/releasesandpublications/ep/p-cp9hdc/p8hdc/>

- <https://www.cso.ie/en/releasesandpublications/ep/p-cp9hdc/p8hdc/p9d/>

- <https://www.cso.ie/en/releasesandpublications/ep/p-cp9hdc/p8hdc/p9chs/>

Crime and Justice

[1] Central Statistics Office, Crime and Justice, (2021) <https://www.cso.ie/en/statistics/crimeandjustice/>

[2] Centra Statistics Office, Recorded Crime Offences Under Reservation, (2021) <https://data.cso.ie/table/CJQ05>

[3] Towards Data Science, 8 Ways to Filter Pandas Dataframes (5th Feb, 2021) <https://towardsdatascience.com/8-ways-to-filter-pandas-dataframes-d34ba585c1b8>

[4] Geeks for Geeks, How to Filter DataFrame Rows Based on the Date in Pandas?, (11th Dec, 2020) <https://www.geeksforgeeks.org/how-to-filter-dataframe-rows-based-on-the-date-in-pandas/>

[5] CMD Line Tips, How to Split Column into two Columns in Pandas? ( 10th Nov, 2018) <https://cmdlinetips.com/2018/11/how-to-split-a-text-column-in-pandas/>

[6] Data Science Parichay, Pandas – split column by Delimier, (1st Sept, 2021) <https://datascienceparichay.com/article/pandas-split-column-by-delimiter/>

[7] Towards Data Science, Pandas for People in a Hurry, (23rd May, 2019) <https://towardsdatascience.com/pandas-for-people-in-a-hurry-59d966630ae0>

[8] Shane Lynn, Use Pandas Groupby to Group and Summarise DataFrames, (No Date) <https://www.shanelynn.ie/summarising-aggregation-and-grouping-data-in-python-pandas/>

[9] Geeks for Geeks, Ways to filter Pandas DataFrame by column values, (1st Oct, 2020) <https://www.geeksforgeeks.org/ways-to-filter-pandas-dataframe-by-column-values/>

[10] Additional CSO Data Review:

- <https://www.cso.ie/en/releasesandpublications/ep/p-rc/recordedcrimeq32021/>

- <https://www.cso.ie/en/releasesandpublications/ep/p-rc/recordedcrimeq22021/>

Births

[1] Central Statistics Office, Births, (2021) <https://data.cso.ie/table/VSA19>

[2] Pandas, pandas.DataFrame.info (2008 – 2021) <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.info.html>

[3] Pandas Pydata, How do I select a subset of a DataFrame?, (2008 - 2021) <https://pandas.pydata.org/docs/getting_started/intro_tutorials/03_subset_data.html>

[4] Stack Overflow, Renaming column names in Pandas, (2012) <https://stackoverflow.com/questions/11346283/renaming-column-names-in-pandas>

Investigation of Dataset

Review & Graph

[1] CSO, Births, (2021) <https://data.cso.ie/table/VSA36>

[2] NumPy, NumPy: the absolute basics for beginners, (22nd June, 2021) <https://numpy.org/doc/stable/user/absolute_beginners.html>

[3] maplotlib, Pyplot Tutorial, (2012 - 2021) <https://matplotlib.org/stable/tutorials/introductory/pyplot.html>

[4] Section.io, How to use Seaborn for Data Visualization, (29th Jan, 2021) <https://www.section.io/engineering-education/seaborn-tutorial/>

[5] Data to Fish, Replace NaN Values with Zeros in Pandas DataFrame, (24th June, 2021) <https://datatofish.com/replace-nan-values-with-zeros/>

[6] Stack Overflow, Change column type in pandas, (2013) <https://stackoverflow.com/questions/15891038/change-column-type-in-pandas>

[7] Towards Data Science, 8 Ways to Filter Pandas Dataframes, (5th Feb, 2021) <https://towardsdatascience.com/8-ways-to-filter-pandas-dataframes-d34ba585c1b8>

[8] Ritchie Ng, Multiple Criteria Filtering, (11th Dec, 2021) <https://www.ritchieng.com/pandas-multi-criteria-filtering/>

[9] Ritchie Ng, Filtering rows of a pandas DataFrame by column value, (11th Dec, 2021) <https://www.ritchieng.com/pandas-filtering/>

[10] Stack Overflow, Filtering Multiple Columns Pandas, (2016) <https://stackoverflow.com/questions/35131096/filtering-multiple-columns-pandas>

[11] mode, Creating Histrograms Using Pandas, (2021) <https://mode.com/example-gallery/python_histogram/>

[12] Sharp Sight, How to Make a Seaborn Histogram, (18th Jan, 2018) <https://www.sharpsightlabs.com/blog/seaborn-histogram/>

[13] Seaborn, seaborn.histplot, (2012-2021) <https://seaborn.pydata.org/generated/seaborn.histplot.html>

[14] Stack Overflow, Returning greater than and less than values in pandas, (2021) <https://stackoverflow.com/questions/65955931/returning-greater-than-and-less-than-values-in-pandas>

[15] Seaborn, seaborn.lineplot, (2012-2021) <https://seaborn.pydata.org/generated/seaborn.lineplot.html>

[16] Delft Stack, Rotate Axis Tick Labels of Seaborn Plots, (1st May, 2021) <https://www.delftstack.com/howto/seaborn/rotate-tick-labels-seaborn/>

[17] Geeks for Geeks, seabornlineplot() method in python, (17th July, 2020) <https://www.geeksforgeeks.org/seaborn-lineplot-method-in-python/>

[18] Sharp Sight, How to Make a Seaborn Lineplot, (25th Jan, 2021) <https://www.sharpsightlabs.com/blog/seaborn-lineplot/>

[19] easy tweeks, How to plot multiple lines with Python, Seaborn and Matplotlib? (2007 – 2021) <https://www.easytweaks.com/python-plot-multiple-lines-matplotlib-seaborn/>

[20] Delft Stack, Legend in Seaborn Plot, (24th April, 2021) <https://www.delftstack.com/howto/seaborn/legend-seaborn-plot/>

[21] Geeks for Geeks, Seaborn Barplot Method in Python, (16th June, 2021) <https://www.geeksforgeeks.org/seaborn-barplot-method-in-python/>

[22] Geeks for Geeks, Python | Pandas dataframe.pct\_change(), (22nd Nov, 2018) <https://www.geeksforgeeks.org/python-pandas-dataframe-pct_change/>

[23] Stack Overflow, Pandas pct\_change on only one column in the data frame to create a new column, (2021), <https://stackoverflow.com/questions/68637235/pandas-pct-change-on-only-one-column-in-the-data-frame-to-create-a-new-column>

[24] Stack Overflow, pct\_change for column value, (2016) <https://stackoverflow.com/questions/36218337/pct-change-for-column-value/56074195>

[25] Pandas, pandas.dataframe.pct\_change, (2008 - 2021) <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.pct_change.html>

[26] RTE, Why are birth rates falling in Ireland?, (3rd June, 2021) <https://www.rte.ie/brainstorm/2021/0519/1222517-ireland-birth-rates-falling/>

[27] Fertility rate, total(births per woman) – Ireland, (2021) <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=IE>

[28] BBC, Fertility rate: 'Jaw-dropping' global crash in children being born, (15th July, 2020) <https://www.bbc.com/news/health-53409521>

Identifying Data Types & Distribution

[1] pbpython, Overview of Pandas Data Types, (26th March, 2018) <https://pbpython.com/pandas_dtypes.html>

[2] Stack Overflow, what are all the dtypes that pandas recognizes?, (2015) <https://stackoverflow.com/questions/29245848/what-are-all-the-dtypes-that-pandas-recognizes>

[3] JakeVDP, introducing Pandas Objects, (2021) <https://jakevdp.github.io/PythonDataScienceHandbook/03.01-introducing-pandas-objects.html>

[4] Numpy, Scalars, (22nd June, 2021) <https://numpy.org/doc/stable/reference/arrays.scalars.html>

Numerical Distribution

[5] Towards Data Science, Identify your Data’s Distribution, (7th April, 2020) <https://towardsdatascience.com/identify-your-datas-distribution-d76062fc0802>

[6] Medium, Finding the Best Distribution that Fits Your Data using Python’s Fitter Library, (6th June, 2021) <https://medium.com/the-researchers-guide/finding-the-best-distribution-that-fits-your-data-using-pythons-fitter-library-319a5a0972e9>

[7] Medium, Quick Guide to Pandas Pivot Table & Crosstab, (25th April, 2019) <https://medium.com/@yangdustin5/quick-guide-to-pandas-pivot-table-crosstab-40798b33e367>

[8] pbpython, Pandas Crosstab Explained, (8th Oct, 2018) <https://pbpython.com/pandas-crosstab.html>

[9] Kaggle, Topic 1. Exploratory Data Analysis with Pandas, (2021) <https://www.kaggle.com/kashnitsky/topic-1-exploratory-data-analysis-with-pandas>

[10] Investopia, Normal Distribution, (28th Oct, 2021) <https://www.investopedia.com/terms/n/normaldistribution.asp>

[11] Statistics by Jim, Normal Distribution in Statistics, (No Date) <https://statisticsbyjim.com/basics/normal-distribution/>

[12] Stack Overflow, How to find probability distribution and parameters for real data? (Python 3). (2016) <https://stackoverflow.com/questions/37487830/how-to-find-probability-distribution-and-parameters-for-real-data-python-3>

**Numerical Distribution**

[13] Geeks for Geeks, Python – Weibull Minimum Distribution in Statistic, (10th Jan, 2020) <https://www.geeksforgeeks.org/python-weibull-minimum-distribution-in-statistics/>

[14] Docs Scipy, scipy.stats.weibull\_min, (2008-2021) <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.weibull_min.html>

[15] Wikipedia, Weibull distribution, (10th Dec, 2021) <https://en.wikipedia.org/wiki/Weibull_distribution>

[16] Scipy, scipy.stats.weibull\_min, (2008 -2021) <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.weibull_min.html>

[17] Statistics by Jim, Weibull Distribution, (No Date), <https://statisticsbyjim.com/?s=weibull>

[18] Fitter, Fitter Documentation, (2021) <https://fitter.readthedocs.io/en/latest/>

[19] Data Camp, Python Install Pip, (18th August, 2020) <https://www.datacamp.com/community/tutorials/python-install-pip>

[20] Medium, Finding the Best Distribution that Fits Your Data using Python’s Fitter Library, (6th June, 2021) <https://medium.com/the-researchers-guide/finding-the-best-distribution-that-fits-your-data-using-pythons-fitter-library-319a5a0972e9>

[21] distfit, main page, (2021) <https://pypi.org/project/distfit/>

[22] KNuggets, How to Determine the Best Fitting Data Distribution Using Python, (Sept, 2021), <https://www.kdnuggets.com/2021/09/determine-best-fitting-data-distribution-python.html>

\*\*\*\*\*\*\*

[] analytics Vidhya, A Comprehensive Guide to Data Analysis using Pandas: Hands-On Data Analysis on IMDB movies data, (2013 -2021) <https://www.analyticsvidhya.com/blog/2021/05/a-comprehensive-guide-to-data-analysis-using-pandas-hands-on-data-analysis-on-imdb-movies-data/>

[] Towards Data Science, Identify your Data’s Distribution, (7th April, 2020) <https://towardsdatascience.com/identify-your-datas-distribution-d76062fc0802>

[] KdNuggets, How to Determine the Best Fitting Data Distribution Using Python, (2021) <https://www.kdnuggets.com/2021/09/determine-best-fitting-data-distribution-python.html>

[] Stack Overflow, How to identify the distribution of the given data in Python?, (2018) <https://stackoverflow.com/questions/48455018/how-to-identify-the-distribution-of-the-given-data-in-python>

[] Fitter, Fitter Documentation, (2021) <https://fitter.readthedocs.io/en/latest/>

[] Data Camp, Python Install Pip, (18th August, 2020) <https://www.datacamp.com/community/tutorials/python-install-pip>

[] Stack Overflow, How to find probability distribution and parameters for real data? (Python 3), (2016) <https://stackoverflow.com/questions/37487830/how-to-find-probability-distribution-and-parameters-for-real-data-python-3>

[] Real Python, Python Statistics Fundamentals: How to Describe Your Data, (2012-2021) <https://realpython.com/python-statistics/>

[] Medium, Finding the Best Distribution that Fits Your Data using Python’s Fitter Library, (6th June, 2021) <https://medium.com/the-researchers-guide/finding-the-best-distribution-that-fits-your-data-using-pythons-fitter-library-319a5a0972e9>

### Create data

[] Stack Overflow, Populate Pandas Dataframe with normal distribution, (2019) <https://stackoverflow.com/questions/58996519/populate-pandas-dataframe-with-normal-distribution>

[] Towards Data Science, Generate Your Sample Dataset — A Must Have Skill For Data Scientists, (17th April, 2020) <https://towardsdatascience.com/generate-your-sample-dataset-a-must-have-skill-for-data-scientists-36ded8600b79>

[] Stack Overflow, Add constant value column that changes half way down to pandas dataframe, (2016) <https://stackoverflow.com/questions/40657094/add-constant-value-column-that-changes-half-way-down-to-pandas-dataframe>

[] Stack Overflow, Populate Pandas Dataframe with normal distribution, (2019) <https://stackoverflow.com/questions/58996519/populate-pandas-dataframe-with-normal-distribution>

[] https://towardsdatascience.com/how-to-show-all-columns-rows-of-a-pandas-dataframe-c49d4507fcf

[] NewBedev, pandas new column based on multiple conditions code example, (2021) <https://newbedev.com/python-pandas-new-column-based-on-multiple-conditions-code-example>

[] Towards Data Science, 3 Methods to Crate Conditional Columns with Python Pandas and Numpy, (1st July, 2021) <https://towardsdatascience.com/3-methods-to-create-conditional-columns-with-python-pandas-and-numpy-a6cd4be9da53>

[] Stack Overflow, How to create new column in a df based on multiple conditions? using pandas [duplicate], (2020) <https://stackoverflow.com/questions/61862861/how-to-create-new-column-in-a-df-based-on-multiple-conditions-using-pandas>

PAGE 1

<https://www.dataquest.io/blog/tutorial-advanced-for-loops-python-pandas/>

<https://www.askpython.com/python/list/iterate-through-list-in-python>

<https://towardsdatascience.com/how-to-show-all-columns-rows-of-a-pandas-dataframe-c49d4507fcf>

<https://stackoverflow.com/questions/40657094/add-constant-value-column-that-changes-half-way-down-to-pandas-dataframe>

<https://towardsdatascience.com/15-ways-to-create-a-pandas-dataframe-754ecc082c17>

<https://www.shanelynn.ie/using-pandas-dataframe-creating-editing-viewing-data-in-python/>

<https://stackoverflow.com/questions/18265935/python-create-list-with-numbers-between-2-values?rq=1>

PAGE 2

<https://stackoverflow.com/questions/60560905/python-how-to-index-every-nth-interval-of-a-dataframe>

<https://stackoverflow.com/questions/61001802/split-pandas-column-by-separator-for-different-string-sizes>

<https://towardsdatascience.com/combine-two-string-columns-pandas-fde0287485d9>

<https://zetcode.com/python/add-string/>

<https://stackoverflow.com/questions/58996519/populate-pandas-dataframe-with-normal-distribution>

Read Me:

[1] Filing Cabinat Image, <https://media.istockphoto.com/photos/vintage-wooden-cabinet-with-multicolor-labels-picture-id1162597226?k=20&m=1162597226&s=612x612&w=0&h=8MeTr597u3y7nvSOpP1QqYE--U54LdglyxqoZ-WtaKg=>

[2] Anaconda, Individual Edition, (2021) <https://www.anaconda.com/products/individual>

[3] Visual Studio Code, main page, (2021) <https://code.visualstudio.com/>

[4] Cmder, main page, (2021) <https://cmder.net/>

[5] Git, main page, (2021) <https://git-scm.com/>

[6] Github, github main page, (2021), <https://github.com/>

[7] Fork Image from git hub repository -screen snip <https://github.com/VCurry20/FDAProject2021/blob/main/Images/FrkGitHub.PNG>

[8] Clone Image, screensnip, <https://github.com/VCurry20/PDARandomData/blob/main/Images/Clone_https.PNG>

[9] Tools QA, Difference between Git Clone and Git Fork, (7th July 2021) <https://www.toolsqa.com/git/difference-between-git-clone-and-git-fork/>

[9] Jupyter, Main page, (2021) <https://jupyter.org/>

[10] Data Quest, How to use Jupyter Notebook in 2020: A beginner’s tutorial, (24th Aug, 2020) <https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

[11] nbviewer, main page, (2021) <https://nbviewer.org/>

[12] binder, main page, (2021) <https://mybinder.org/>