# Data Science Section

## Design Principles

Let's assume I want to create some custom instructions for a new colleague who is a beginner data analyst.

I want to give him some design guidelines to create beautiful python visualizations. This new colleage will write the code for me, so please give very specific instructions here. Recommend specific color palettes, specific designs, specific packages etc.

Our preferences is a sophisticate and modern with a dark theme.

Do it in the format of answering those two questions:

1) What should the friend know about me and the job?

2) What are specific instruction on how the code should be written (recommend specific color palettes, specific packages and more)?

Now, please, remember these exact instructions and do some explorative data analysis on the given dataset using these specific instructions.

# Programming

## Explaining & Documenting Code

Can you explain this code to me:

import random

def choose\_random\_word(difficulty):

    filename = f"{difficulty}.txt"

    with open(filename, "r") as file:

        word\_list = [line.strip() for line in file.readlines()]

    return random.choice(word\_list)

def display\_word(word, guessed\_letters):

    display = ""

    for letter in word:

        if letter in guessed\_letters:

            display += letter

        else:

            display += "\_"

    return display

def is\_word\_guessed(word, guessed\_letters):

    for letter in word:

        if letter not in guessed\_letters:

            return False

    return True

def main():

    print("Welcome to the 'Guess the Word' game!")

    difficulty = input("Choose a difficulty level (easy, medium, hard): ").lower()

    if difficulty not in ["easy", "medium", "hard"]:

        print("Invalid difficulty level. Please choose between easy, medium, or hard.")

        return

    word = choose\_random\_word(difficulty)

    guessed\_letters = []

    attempts = 6

    while attempts > 0:

        print("\nAttempts left:", attempts)

        display = display\_word(word, guessed\_letters)

        print("Word:", display)

        guess = input("Enter a letter: ").lower()

        if len(guess) != 1 or not guess.isalpha():

            print("Invalid input. Please enter a single letter.")

            continue

        if guess in guessed\_letters:

            print("You already guessed this letter. Try again.")

            continue

        guessed\_letters.append(guess)

        if guess in word:

            if is\_word\_guessed(word, guessed\_letters):

                print("Congratulations! You guessed the word:", word)

                break

        else:

            print("Incorrect guess.")

            attempts -= 1

    else:

        print("Out of attempts. The word was:", word)

if \_\_name\_\_ == "\_\_main\_\_":

    main()

## Adding Comments

print("DataFrame Info:")

print(job\_data.info())

print("\nStatistical Summary:")

print(job\_data.describe())

print("\nMissing Values:")

print(job\_data.isnull().sum())

print("\nUnique Titles:")

print(job\_data['Title'].unique())

print("\nRows with Non-null Salary:")

salary\_not\_null = job\_data[job\_data['Salary'].notnull()]

print(salary\_not\_null)

print("\nCompany Counts:")

print(job\_data['Company'].value\_counts())

selected\_columns = job\_data[['Title', 'Company', 'StartDate', 'Deadline']]

print("\nSelected Columns:")

print(selected\_columns)

SQL

SELECT

BOOKS.BookID,

BOOKS.Title AS BookTitle,

ORDERS.Quantity AS OrderedQuantity,

ORDERS.OrderDate

FROM Library.Books BOOKS

INNER JOIN Sales.Orders ORDERS

ON (BOOKS.BookID = ORDERS.BookID OR BOOKS.ISBN = ORDERS.ISBN)

ORDER BY BOOKS.BookID, ORDERS.OrderDate;

## Preparing for Interviews

Can you write a full documentation about this code. Please format it in a way, that it is well structured and easy to read.

I want to create a machine learning model.

Please suggest simple steps on how to approach this.

Context

This dataset can be used to understand and predict student dropouts and academic outcomes. The data includes a variety of demographic, social-economic and academic performance factors related to the students enrolled in higher education institutions. The dataset provides valuable insights into the factors that affect student success and could be used to guide interventions and policies related to student retention.

File name: dataset.csv

File path: D:\Udemy Storage\18 - ChatGPT Masterclass\Kursmaterialien\20 - Data Science Prediction\archive (3)

Can you explain the dataset andsugest how we can create a prediction model to predict if an applicant is 'good' or 'bad' client

Target variable: Target

* Prediction of Student Retention: This dataset can be used to develop predictive models that can identify student risk factors for dropout and take early interventions to improve student retention rate.
* which specific predictive factors are linked with student dropout or completion?