# **BWF-Week 1**

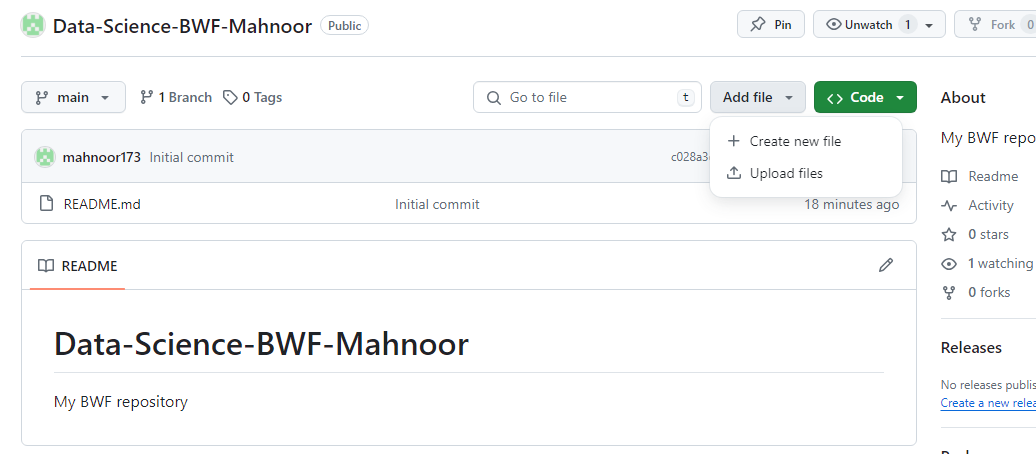
# **Task1:**

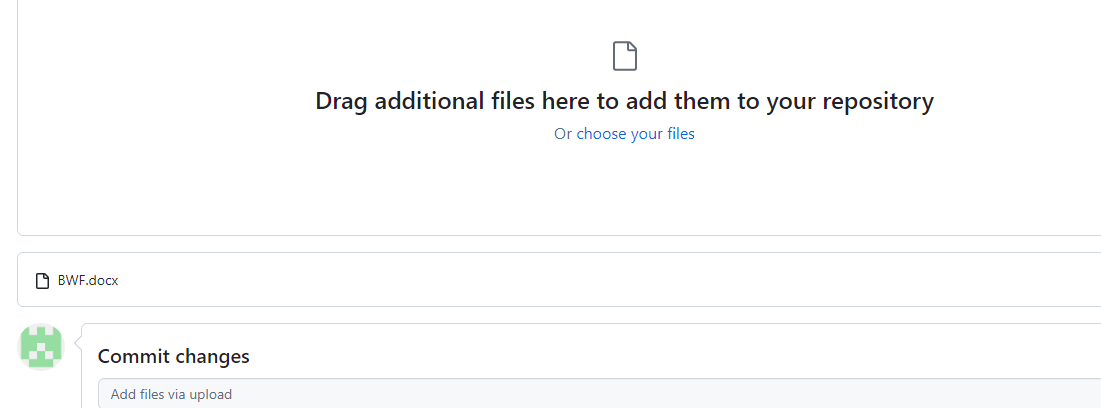
**Understanding Data Science:**

Data Science is a scientific field that uses scientific methods to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Data science uses scientific methods, such as probability and statistics, and operates on both structured and unstructured data. Data science studies how data can be gathered, stored and operated upon using computers. Data science is a broad field, touching many other disciplines, including databases, big data, machine learning, artificial intelligence, visualization, and more. Data ethics is a new branch of ethics that studies and evaluates moral problems related to data, algorithms and corresponding practices. Data ethics concepts include shared values (principles) and ethical challenges (problems) for data ethics, and case studies that help understand these concepts in real-world contexts. Data ethics principles include accountability, transparency, fairness, reliability & safety, privacy & security, and inclusiveness. Data ethics challenges include data ownership, informed consent, intellectual property, data privacy, the right to be forgotten, dataset bias, data quality, algorithm fairness, misrepresentation, and the illusion of free choice. Data ethics culture is about operationalizing applied ethics to make sure that ethical principles and practices are adopted in a consistent and scalable manner across the entire organization. Data can be described as raw, quantitative, or qualitative. Data can be structured, unstructured, or semi-structured. Data sources include databases, files, internet sources, application programming interfaces, and web scraping. Probability is a number between 0 and 1 that expresses how probable an event is. Random variables are used to describe events, and probability distribution describes the probability of a random variable taking a certain value. Mean, variance, and standard deviation are used to describe the distribution of a random variable. Confidence intervals are used to estimate the true mean of a population given a sample. Hypothesis testing is used to test whether a hypothesis about a population is true. The law of large numbers and the central limit theorem are important concepts in probability and statistics. Covariance and correlation are used to describe the relationship between two random variables.

# **Task2:**

In Task 2, I made my github account and created my reporitory names Data-Science-BWF-Mahnoor for uploading my tasks.





# **Task3:**

The three measures of central tendency are the mean, median, and mode. They are used to describe a set of data with a smaller set of numbers.

* **Mean:** The sum of all numbers divided by the number of data points. It is a human-constructed definition that is commonly used to represent a typical or middle value.
* **Median:** The middle value in a dataset when it is ordered from smallest to largest. If there are an even number of data points, the median is the average of the two middle values.
* **Mode:** The most common value in a dataset. If there is no single most common value, then there is no mode.