ARIF ALAM

# 100 DAYS OF DATA SCIENCE

A guide to start your data science journey

# The Author

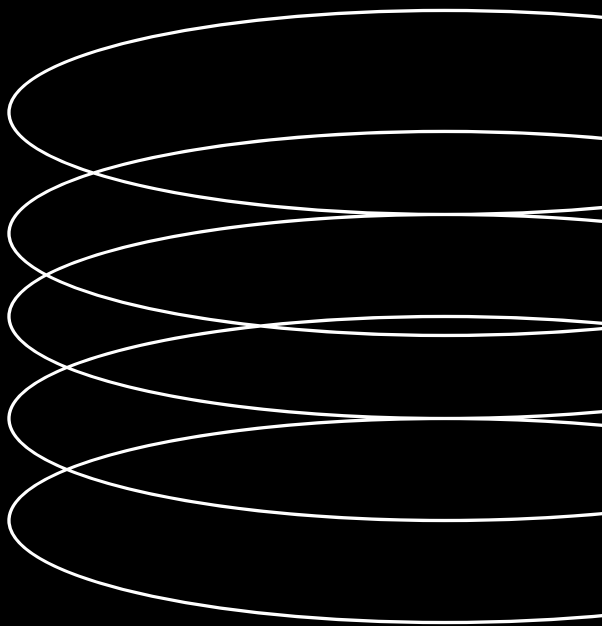
HELLO THERE! I'M ARIF

As a software engineer and author, I specialize in creating technical content that helps developers and IT professionals stay up-to-date on the latest technologies and best practices. Whether I'm writing about programming languages, software development methodologies, or data science, I aim to provide clear, actionable guidance that my readers can use to improve their skills and advance their careers.

.

# Lets start your

# Data science Journey



A STEP-BY-STEP GUIDE

The aim of this 100 Days series is to get you started assuming that you have no prior knowledge of any of these topics.

In this series we will try to keep everything using Python, less because of its robustness but more because the codes in Python are self explanatory.

You are expected to give anywhere between
2-3 hours per day.

Don’t jump to the next topic until and unless you are done with the previous day syllabus.

CHAPTER N.1

# Day 1 - 10 Python Basics

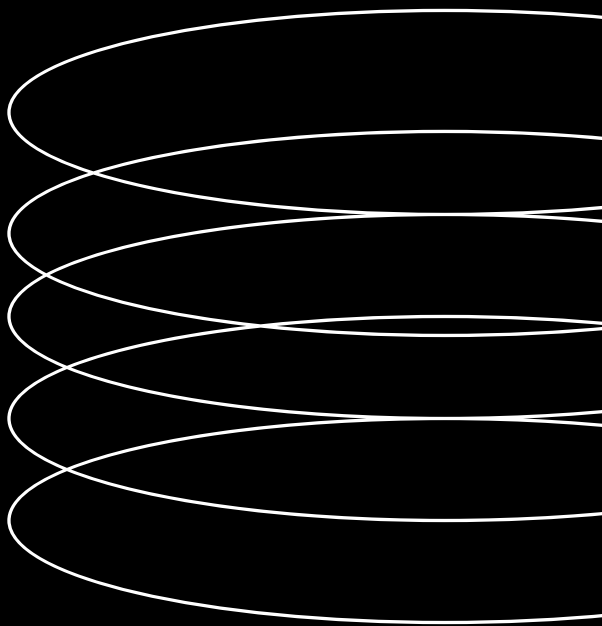
## Here are some things you could do during the first 10 days to get started with the basics:

- Day 1: Set up a Python development environment. This could involve installing Python on your computer, setting up a virtual environment, and installing any libraries or packages you will be using.
- Day 2: Familiarize yourself with the basic syntax of Python. This could include learning about variables, data types, operators, and control structures like loops and conditional statements.
- Day 3: Practice using Python to work with data. You could start by loading and manipulating simple datasets using Python's built-in data structures and functions.
- Day 4: Learn about Python's object-oriented programming features. This could include learning about classes, inheritance, and polymorphism.
- Day 5: Practice using Python to perform basic data analysis tasks. This could include tasks like calculating summary statistics, generating plots, and performing simple statistical tests.
- Day 6: Start working on a small Python project or problem. This could be something simple, like building a basic command-line tool or web scraper.

- Day 7: Learn about Python's support for working with databases. This could include learning how to connect to databases, execute SQL queries, and work with data stored in a database.
- Day 8: Learn about Python's support for working with data in the cloud. This could include learning about tools like Amazon Web Services (AWS) or Google Cloud Platform (GCP) and how to use Python to access cloud-based data and services.
- Day 9: Practice working with data using Python's scientific computing libraries, such as NumPy and Pandas.
- Day 10: Join an online community or find a study group to connect with others who are learning Python and data science. This could be a great way to get support and stay motivated as you continue your journey.

“When we have all data online it will be great for humanity. It is a prerequisite to solving many problems that humankind faces.”

~ ROBERT CAILLIAU



CHAPTER N.2

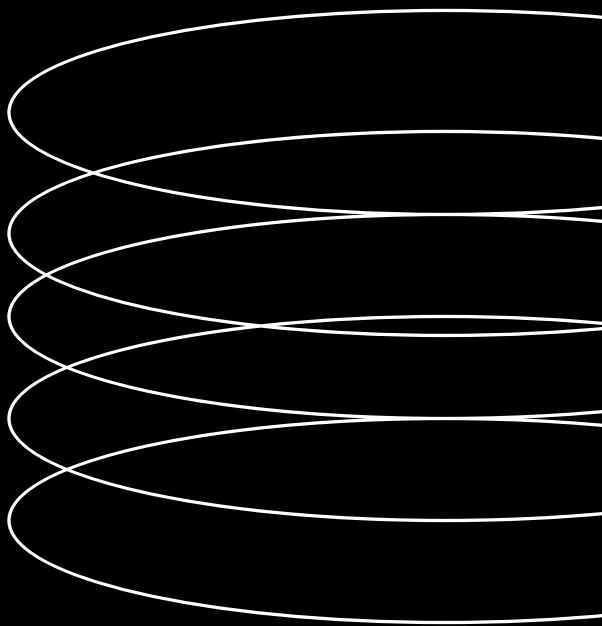
# Day 10-20 Data Types

## Here is a suggested plan for learning about data types for data science during days 10-20 of a 100 days of data science challenge:

- Day 10: Begin by familiarizing yourself with the basics of data types in your chosen programming language, such as integers, floats, and strings. Work through online tutorials and practice problems to gain a solid understanding of these concepts.
- Day 11-13: Continue learning about data types, and explore more advanced concepts such as boolean values, arrays, and lists. Practice creating and manipulating variables of different data types.
- Day 14-16: Learn about data structures, such as dictionaries, sets, and tuples, and practice using these structures to store and manipulate data.
- Day 17-18: Practice working with data types and data structures by applying them to real-world data sets. This could involve loading and cleaning data, calculating statistics, or creating visualisations
- Day 19-20: Explore machine learning concepts and techniques, and consider how data types and data structures can be used in the machine learning process. Practice applying these concepts to simple data sets.

“Data are just summaries of thousands of stories—tell a few of those stories to help make the data meaningful.”

~ DAN HEATH,
BESTSELLING AUTHOR



CHAPTER N.3

# Day 20-30 Statistics and Calculus

## Here is a suggested plan for learning about statistics and calculus for data science during days 20-30 of a 100 days of data science challenge:

- Day 20: Begin by familiarizing yourself with the basics of statistics, such as mean, median, and standard deviation. Work through online tutorials and practice problems to gain a solid understanding of these concepts.

- Day 21-23: Continue learning about statistics, and explore more advanced concepts such as probability, hypothesis testing, and regression analysis. Practice applying these concepts to real-world data sets.

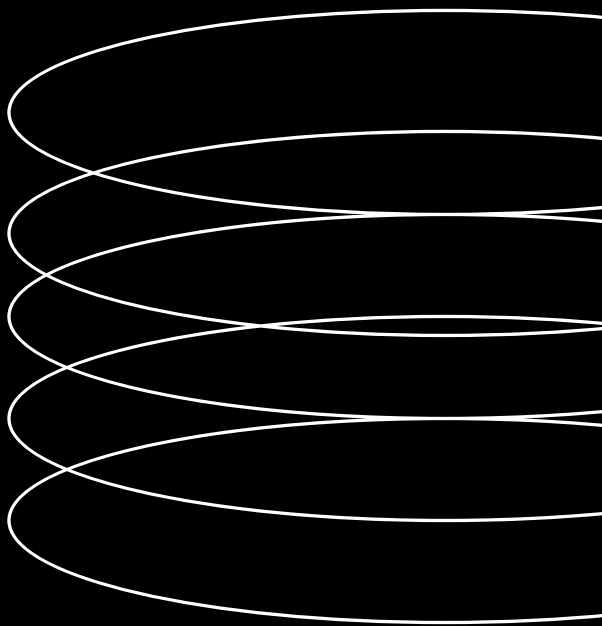
- Day 24-26: Learn about calculus, including concepts such as derivatives, integrals, and optimization. Practice solving calculus problems and understand how these concepts can be applied in data science.

- Day 27-29: Practice applying your knowledge of statistics and calculus to real-world data sets, and consider how these concepts can be used in machine learning.

- Day 30: Reflect on your progress and set new goals for the future. Consider seeking out additional resources or challenges to continue learning and growing as a data scientist.

“Person who is better at statistics than any software engineer and better at software engineering than any statistician.”

~ JOSH WILLS



CHAPTER N.4

# Day 30-40 Basic Machine Learning

# Here is a suggested plan for learning about machine learning for data science during days 30-40 of a 100 days of data science challenge:

Day 30: Begin by familiarizing yourself with the basics of machine learning, including concepts such as supervised and unsupervised learning, and common algorithms such as linear regression and k-means clustering. Work through online tutorials and practice problems to gain a solid understanding of these concepts.

Day 31-33: Continue learning about machine learning, and explore more advanced concepts such as decision trees, neural networks, and deep learning. Practice applying these concepts to simple data sets.

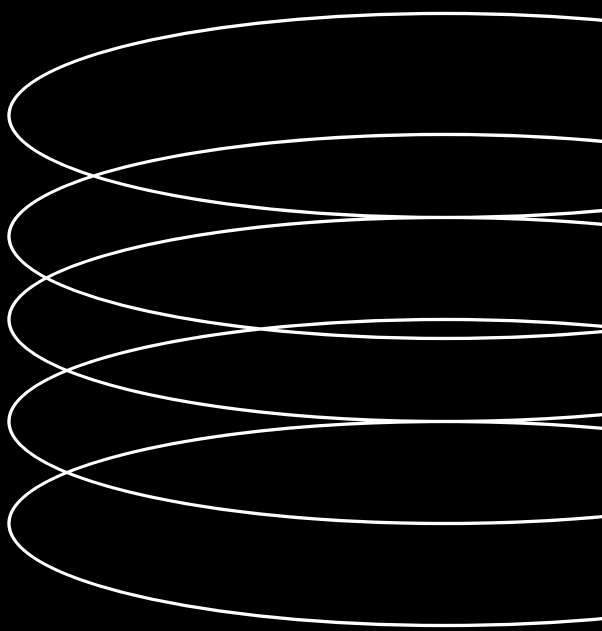
Day 34-36: Learn about common tools and libraries for building machine learning models, such as scikit-learn and TensorFlow. Practice using these tools to build and evaluate models.

Day 37-39: Practice applying your knowledge of machine learning to real-world data sets, and consider how these concepts can be used to solve specific problems or tasks.

Day 40: Reflect on your progress and set new goals for the future. Consider seeking out additional resources or challenges to continue learning and growing as a data scientist.

“The core advantage of data is that it tells you something about the world that you didn’t know before.”

~ HILARY MASON



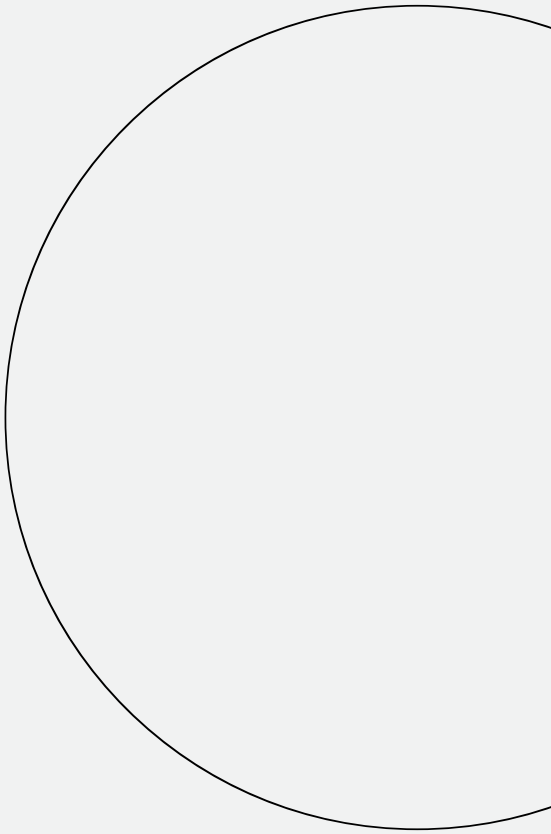
CHAPTER N.5

# Day 40-50 Basic Deep Learning

## Here are some suggestions for what you could focus on during days 40-50 of your 100-day deep learning data science challenge:

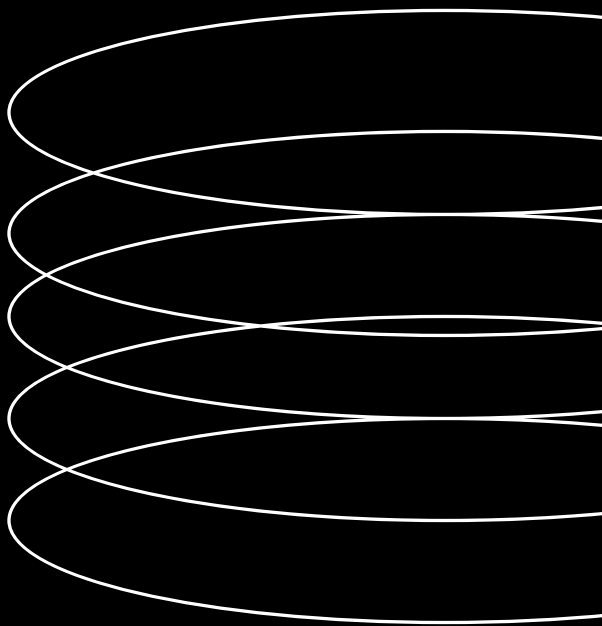
- Day 40: Transfer learning: This is a technique where you can use a pre-trained model on a large dataset and fine-tune it for your own task. This can be a quick and effective way to get good performance on a new problem, especially if you have a small dataset.
- Day 41-43: Hyper-parameter tuning: Once you have a model set up, you will need to fine-tune its hyper-parameters in order to get the best performance. This can be done using techniques such as grid search or random search.
- Day 44-46: Evaluating model performance: It's important to understand how to evaluate the performance of your model. This includes understanding different evaluation metrics (such as accuracy, precision, and recall) and knowing how to plot and interpret learning curves.
- Day 47-49: Regularisation is a technique that can help prevent overfitting and improve the generalisations of your model. There are several types of regularisation, including L1 and L2 regularization, dropout, and early stopping.

- Day 50: Convolutional neural networks (CNNs): CNNs are a type of neural network that are particularly well-suited for image classification tasks. You can learn about the architecture of CNNs and how they work, and then practice building and training CNNs using a framework such as PyTorch or TensorFlow.



“We are surrounded by data, but starved for insights.”

~ JAY BAER



CHAPTER N.6

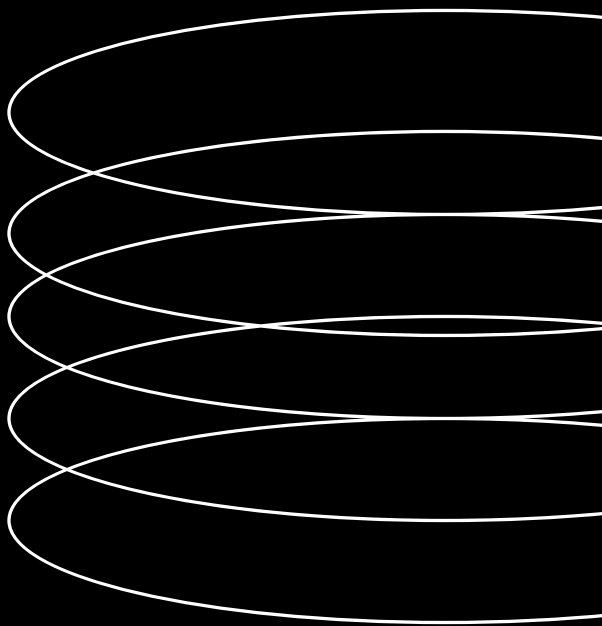
# Day 50-60 Data Visualisation

## Here are some suggestions for what you could focus on during days 50-60 of your 100-day data visualization data science challenge:

- Data visualisation libraries: There are several libraries that you can use to create data visualisations, including Matplotlib, Seaborn, and Plotly. You can learn about the different types of plots that each library offers and practice creating visualizations with them.
- Visual encoding: When creating data visualizations, it's important to choose the right visual encodings (such as position, length, and color) to effectively convey the data. You can learn about different visual encodings and how to use them effectively.
- Visual design principles: There are several principles of visual design that you can use to create effective data visualizations, such as using appropriate chart types and using appropriate scales. You can learn about these principles and practice applying them to your own visualizations.
- Storytelling with data: Data visualization is not just about creating pretty pictures, it's also about telling a story with the data. You can learn about different techniques for storytelling with data and practice creating visualizations that effectively communicate a message.
- Interactive visualizations: You can create interactive visualizations using libraries such as Plotly or Bokeh, which allow users to explore the data in more depth. You can learn about how to create interactive visualizations and practice building them.

“Data are just summaries of thousands of stories.”

~ CHIP & DAN HEATH



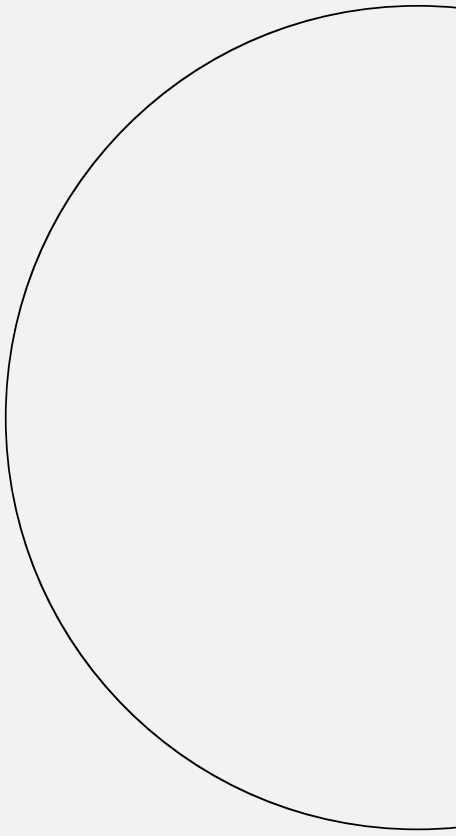
CHAPTER N.7

# Day 60-70 Data Cleaning

## Here are some suggestions for what you could focus on during days 60-70 of your 100-day data cleaning data science challenge:

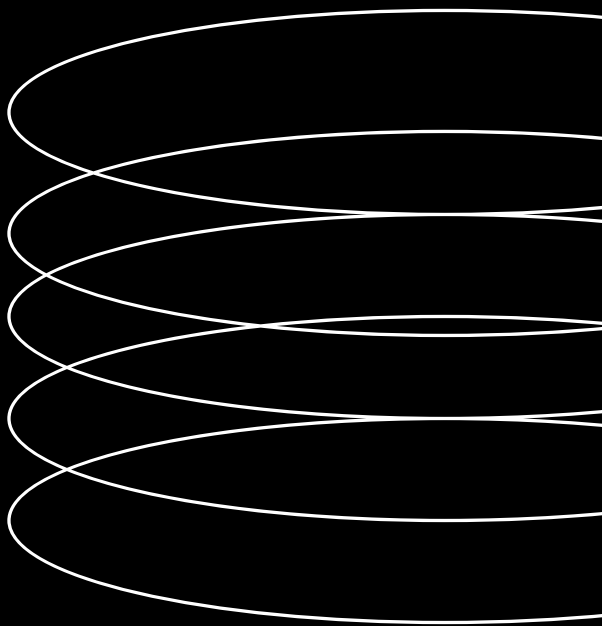
- Identifying and handling missing values: It's important to identify and handle missing values in your data, as they can impact the performance of your model. You can learn about different techniques for handling missing values, such as imputation or dropping the rows with missing values.
- Handling outliers: Outliers can have a significant impact on your data and the results of your analysis. You can learn about different techniques for detecting and handling outliers, such as using statistical methods or visualisation.
- Data transformation: You may need to transform your data in order to prepare it for analysis or modeling. This can include scaling, normalisation, and encoding categorical variables.
- You can learn about different data transformation techniques and practice applying them to your data.

- Data cleaning with Python: There are several Python libraries that can be useful for data cleaning tasks, such as Pandas and Numpy. You can learn about these libraries and how to use them to clean and prepare your data.
- Data cleaning best practices: It's important to follow best practices when cleaning your data, such as keeping a record of your cleaning steps and testing your data after cleaning to ensure it's correct. You can learn about these best practices and how to apply them to your data cleaning process.



“No data is clean, but most is useful.”

~ DEAN ABBOTT,



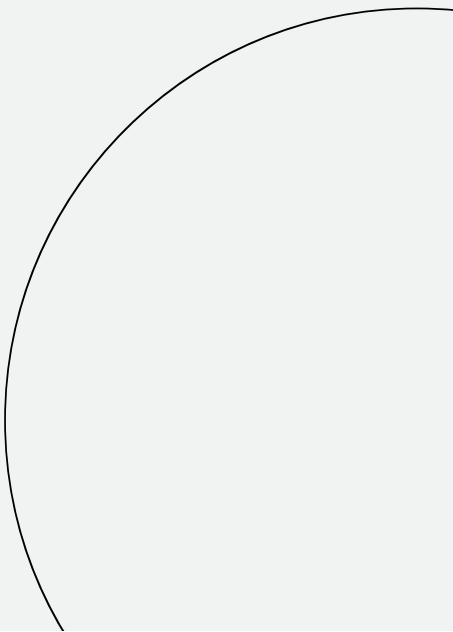
CHAPTER N.8

# Day 70-80 Some Projects

## Here are some ideas for projects that you could work on during the 100 Days of Data Science Challenge:

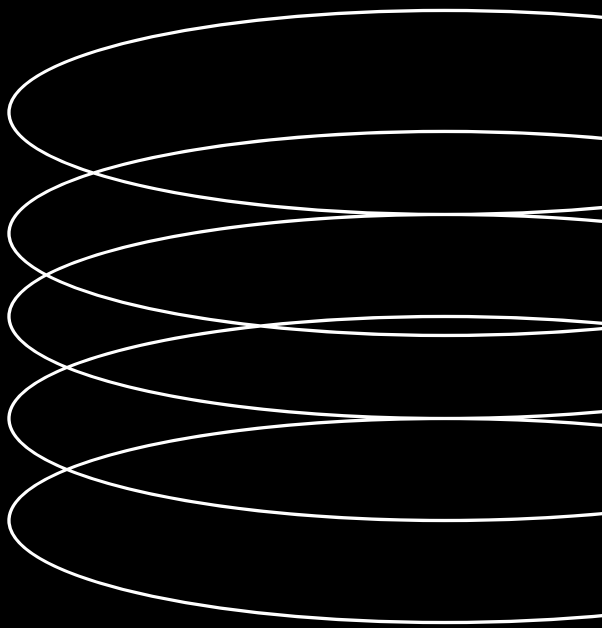
- **Predicting stock prices**: In this project, you could gather data on historical stock prices and try to build a machine learning model that can predict future stock prices. You could try using different features, such as company financial statements, news articles, and market indicators, to see which ones have the most predictive power.
- **Detecting spam emails**: In this project, you could gather a dataset of spam and non-spam emails and build a machine learning model that can classify new emails as spam or not spam. You could try using different features, such as the words in the email, the sender's domain, and the presence of links or attachments, to see which ones are most effective at identifying spam emails.
- **Predicting housing prices**: In this project, you could gather data on housing prices in a specific area and try to build a machine learning model that can predict the sale price of a new house. You could try using different features, such as the size of the house, the number of bedrooms and bathrooms, and the location, to see which ones have the most influence on the sale price.

- **Analysing customer behaviours** : In this project, you could gather data on customer behaviour, such as purchase history and website activity, and build a machine learning model that can predict which customers are most likely to make a purchase. You could try using different features, such as the products that customers have viewed, the time of day that they visit the website, and the frequency of their visits, to see which ones are most predictive of purchase behaviour.
- **Sentiment analysis** : In this project, you could gather a dataset of text data, such as reviews or social media posts, and build a machine learning model that can classify the sentiment of the text as positive, negative, or neutral. You could try using different techniques, such as bag-of-words or word embeddings, to represent the text data and see which ones give the best performance.



“Data is like garbage. You’d better know what you are going to do with it before you collect it.”

~ MARK TWAIN



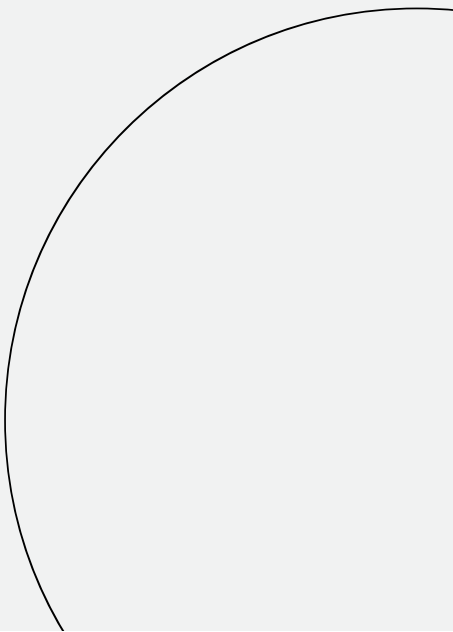
CHAPTER N.9

# Day 80-90 Communication Skills

**Developing good communication skills is an important aspect of the 100 Days of Data Science Challenge, as it can help you to effectively share your findings and insights with others.**

- **Here are some ways that you might work on your communication skills during the challenge:**
- Practice presenting your work: Practice presenting your work to others, whether in person or virtually. This can help you to develop your skills in presenting information clearly and concisely, and in answering questions about your work.
- Write about your work: Consider writing about your work in the form of a blog post or report. This can help you to practice explaining complex concepts in a clear and understandable way, and to organize your thoughts and findings in a logical manner.
- Seek feedback from others: Ask for feedback from others on your communication skills, and use this feedback to identify areas for improvement. This might involve asking for feedback from peers, mentors, or others who have experience in data science.

- Read and learn from others: Read articles, papers, and other materials written by experts in data science, and pay attention to the way that they communicate their ideas and findings. This can help you to develop an understanding of what makes for effective communication in the field.
- By working on your communication skills during the 100 Days of Data Science Challenge, you can improve your ability to share your work and insights with others, which can be beneficial for your future career in data science.



“If you wanna do data science, learn how it is a technical, cultural, economic, and social discipline that has the ability to consolidate and rearrange societal power structures.”

~ HUGO BOWNE-ANDERSON

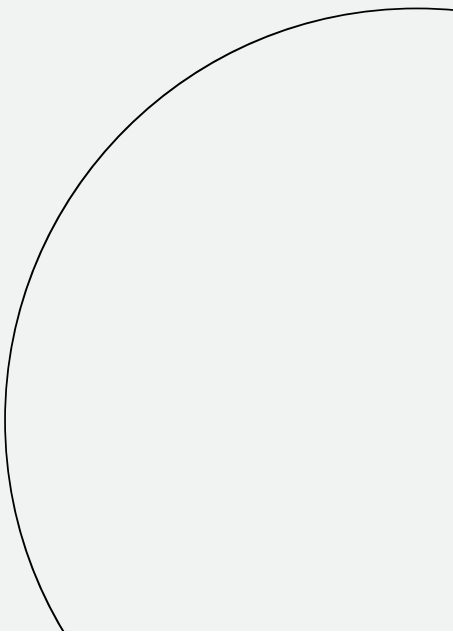
CHAPTER N.9

# Day 90-100 Revision

**It is important to periodically review and revise your progress in the 100 Days of Data Science Challenge, as this can help you to stay motivated and on track, and identify areas where you may need to focus more attention. Here are some suggestions for how you might approach reviewing your progress:**

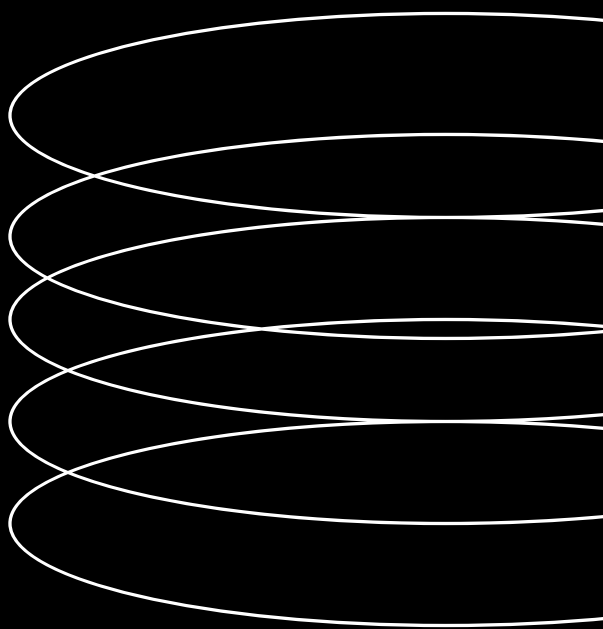
- Review your project goals and objectives: Take some time to review your project goals and objectives and ensure that you are making progress towards achieving them. Consider whether you need to adjust your goals or modify your approach in order to make more progress.
- Reflect on your strengths and challenges: Think about the skills and techniques that you have learned so far, and identify any areas where you feel particularly strong or weak. This can help you to focus your efforts on the areas where you need the most improvement.
- Review your project work so far: Take some time to review the work that you have completed so far on your project. Consider whether you are happy with your progress, or if there are any areas where you feel you could have done better.

- Seek feedback from others: Consider asking for feedback from others, such as your peers or mentors, on your project work so far. This can help you to get a fresh perspective and identify areas for improvement that you might have missed.
- Revise your project plan: Based on what you have learned so far and the feedback that you have received, consider revising your project plan to incorporate any changes or adjustments that you think might be necessary.
- By regularly reviewing and revising your progress in the 100 Days of Data Science Challenge, you can ensure that you are making steady progress towards your goals and staying on track to succeed in your project.



“In the end you should only measure and look at the numbers that drive action, meaning that the data tells you what you should do next.”

~ ALEXANDER PEINIGER



Want more Guide like this?

# Data Science Reality

**FOLLOW FOR MORE**