

Turing Academy

Data Analytics

Course Syllabus

Introduction

Welcome to the Data Analytics Specialization course at Turing Academy. Over the next 6 months, you will embark on an intensive journey to master Data Analytics, followed by an internship program to apply your knowledge in a real-world setting.

The course is structured into 5 modules:

1. **Foundation of Data Analysis with Excel**
2. **Data Visualization and Reporting with PowerBI**
3. **Data Management and Querying with SQL**
4. **Python for Data Analysis and Data Wrangling**
5. **Statistical Methods and Probability Theory (Throughout whole course)**

Each module is designed to build on the previous one, ensuring a progressive and comprehensive understanding of backend development. The learning experience is enhanced through a blend of theoretical knowledge and practical application.

Course Structure and Schedule

Lectures are conducted 2-3 times a week, depending on the module. To reinforce the knowledge gained in lectures, the course includes a variety of supplementary activities:

- **Additional Reading & Watching Materials:** Curated resources to deepen your understanding of each topic.
- **Step Projects:** Collaborative assignments that mimic real-world Data Analytics problems.
- **Home Tasks:** Regular homework to practice and apply new concepts.
- **Lab Days:** Hands-on tests to assess practical skills.
- **Quizzes:** Short assessments to ensure comprehension of key concepts.
- **Midterm and Final Exams:** Comprehensive exams to evaluate your overall understanding and retention.
- **Interviews:** Mock interviews to prepare you for job placements and real-world scenarios.
- **Sprints:** Two-week project work with developers and designers to simulate real-world product development.

Evaluation and Grading

Your performance throughout the course will be assessed based on a weighted grading system:

- ***Labs: 20%***
- ***Step and Final Projects: 30%***
- ***Quizzes: 10%***
- ***Midterm and Final Exams: 30%***
- ***Interviews: 10%***

This grading system is designed to provide a balanced evaluation of your theoretical knowledge, practical skills, collaborative abilities, and interview readiness.

Syllabus for Data Analysis Course

Module 1: Foundations of Data Analysis with Excel

Data Management and Formatting

- **Worksheet Organization:** Efficiently navigating and managing multiple sheets.
- **Data Sorting & Filtering:** Using basic and custom sorting, advanced filter options.
- **Data Entry Automation:** Applying Flash Fill, advanced Fill Series options.
- **Custom Formatting:**
 - Number Formatting: Creating custom formats for numerical data.
 - Date and Time Formatting: Custom formats for time-sensitive data.
- **Data Validation and Control:** Setting up data validation rules for cleaner inputs.
- **Grouping and Outlining:** Grouping, outlining, and summarizing large datasets.
- **Subtotal and Totals Calculation:** Adding summaries with subtotals and totals.
- **Duplicate Management:** Identifying and removing duplicate data points.
- **Watch Window:** Monitoring key data points in large spreadsheets.
- **Hyperlinks:** Creating links to external sources or cells within the workbook.
- **Protection and Security:** Protecting sheets and entire workbooks.

Advanced Formula Auditing and Troubleshooting

- **Tracing Dependencies and Precedents:** Understanding relationships between cells.
- **Error Checking and Formula Auditing:**
 - Formula Watch and Error Checking.
 - Using Show Formulas, Evaluate Formula tools for troubleshooting.
- **Data Transformation Tools:**
 - Text to Columns: Splitting data by delimiter.
 - Consolidate: Merging data from different sheets.
- **Name Manager:** Organizing named ranges for easier formula management.

Conditional Formatting and Data Visualization

- **Conditional Formatting Rules:**
 - Predefined and Custom Formatting Rules.
 - Using Formulas in Conditional Formatting.
- **Table Formatting:** Converting data ranges into tables with unique styles.
- **Text Wrapping and Alignment:** Managing large texts with Wrap Text and Merge & Center.
- **Freezing Panes:** Locking headers for large datasets.
- **Comments and Notes:** Adding, editing, and managing comments for clarity.

Advanced Charting and Interactive Dashboards

- **PivotTables and Data Grouping:**
 - Creating and modifying PivotTables.
 - Grouping data for summarization.
- **Pivot Charts and Slicers:**
 - Customizing slicers for better dashboard filtering.
 - Using PivotCharts to visualize PivotTable data.
- **Advanced Charting Techniques:**
 - Creating interactive and custom charts (waterfall, funnel, gauge charts).
 - Applying styles and formatting to charts.
- **Dashboard Design Principles:**
 - Fundamentals of effective dashboard design.
 - Creating and modifying interactive dashboards.
- **Timeline and Data Refresh Options:**
 - Timeline Creation: Adding date-driven filters.
 - Refresh and Data Source Management: Maintaining up-to-date dashboards.

Data Analytics with Excel Formulas (100+ Formulas)

- **Lookup & Reference Functions:**
 - VLOOKUP, HLOOKUP, INDEX, MATCH, CHOOSE, SWITCH, GETPIVOTDATA.
- **Statistical Functions:**
 - COUNT, COUNTA, COUNTIF, COUNTIFS, AVERAGE, AVERAGEIF, MAX, MIN, STDEV.
- **Mathematical Functions:**
 - SUM, SUMIF, SUMPRODUCT, ROUND, ABS, MOD, QUOTIENT, POWER.
- **Text Functions:**
 - LEFT, RIGHT, MID, CONCATENATE, FIND, REPLACE, TRIM, TEXT.
- **Date & Time Functions:**
 - YEAR, MONTH, DAY, NETWORKDAYS, WORKDAY, WEEKNUM, TODAY, NOW.
- **Logical Functions:**
 - IF, AND, OR, NOT, IFERROR, IFNA.
- **Financial Functions:**
 - PMT, NPER, FV, PV.
- **Array Formulas and Dynamic Arrays:**
 - Understanding array formulas and functions like UNIQUE, FILTER, SORT.

Data Cleaning and Preparation

- **Column and Row Reorganization:** Moving, inserting, and deleting rows and columns.
- **Text-to-Number Conversion:** Converting text-formatted numbers for calculations.
- **Advanced Column Operations:**
 - Splitting data using Text to Columns.
 - Auto-filling missing values and detecting blank cells.
- **Automated Data Cleaning:**
 - Techniques for identifying and removing errors, blanks, and inconsistencies.
- **Text Manipulation for Cleaning:**
 - Techniques for cleaning text data (TRIM, UPPER/LOWER, CLEAN).

Basic Statistical Analysis with Excel

- **Descriptive Statistics:**
 - Calculation of mean, median, mode, range, and quartiles.
 - Application of standard deviation and variance.
- **Correlation Analysis:** Analyzing relationships using correlation functions.
- **Outlier Detection and Boxplots:**
 - Identifying outliers.
 - Visualizing distributions with Boxplot charts.
- **Frequency Distributions:** Calculating frequency and cumulative distributions.

Power Query for Data Automation

- **Data Connection and Importing:** Connecting to different data sources.
- **Data Merging and Transformation:**
 - Combining multiple files or tables.
 - Automating repetitive data cleaning tasks.
- **Custom Queries:** Creating and modifying custom queries.

Introduction to VBA (Visual Basic for Applications)

- **Developer Tab:** Enabling and exploring the Developer environment.
- **Macro Basics:**
 - Recording basic macros.
 - Assigning macros to buttons.
- **Basic VBA Code Writing:**
 - Writing initial codes and adding comments.
 - Reading and editing recorded macros.
- **Final Project:** Applying Excel skills to deliver a complete data analysis project.

Module 2: Data Visualization and Reporting with Power BI Desktop and Power Query

Introduction to Power BI Desktop and Power Query

- **Familiarity with Visualization Tools:** Overview of popular data visualization tools.
- **Introduction to Power BI:** Understanding the purpose, capabilities, and real-world applications.
- **Power BI Workflow:**
 - Downloading and installing Power BI Desktop.
 - Exploring workflow and data handling processes in Power BI.
- **General Interface Overview:** Navigating the main components and features of the Power BI interface.
- **Loading Initial Data Sources:** Connecting to and loading datasets for analysis.
- **Data Modeling Basics:** Introduction to data modeling concepts within Power BI.
- **Data Review and Exploration:** Basic data analysis and exploration skills in Power BI.
- **Using the Power Query Editor:**
 - Overview of Power Query and its application in data transformation.
- **Filtering Options:**
 - Text and Number Filters.
- **Row and Column Management:**
 - Techniques for managing and organizing data structure.
- **Reference and Duplicate Creation:**
 - Creating references and duplications on demand.
- **Data Type Transformation:** Setting and changing data types.
- **Value Substitution:** Replacing or updating values.
- **Managing Headers:** Adjusting header rows and formatting.
- **Column Splitting:** Using “Split Column” to manage data fields.
- **Data Source Management:** Connecting, updating, and refreshing data sources.
- **Group By Function:** Grouping data for summarization and aggregation.

Advanced Power Query Operations

- **Merging Queries:** Combining data tables using “Merge Queries.”
- **Appending Queries:** Consolidating datasets using Append Queries.
- **Pivot and Unpivot Columns:** Transforming data structure for analysis.
- **Specifying Line Counts:** Setting row count parameters for data handling.
- **Primary Mathematical Operations:** Using Power Query for basic mathematical transformations.
- **Preliminary Data Analysis:** Techniques for reviewing initial datasets.
- **Data Indexing:** Adding indexes to data for improved referencing.
- **Column Management:**
 - Adding new columns, updating data, and formatting transformations.

Data Modeling and Report View

- **Understanding Model Relationships:** Establishing and visualizing relationships between tables.
- **Star Schema:** Managing relationships in a star schema structure.
- **Fact vs. Dimension Tables:** Differences and significance in data modeling.
- **Data Visualization Principles:** Key practices for effective and informative visualizations.
- **Using Tooltips:** Enhancing report interactivity with tooltips.
- **Slicers:** Filtering data dynamically within reports.
- **Synchronization Principles:** Ensuring data consistency across visualizations.
- **Graph and Chart Operations:** Creating, formatting, and customizing charts and graphs.
- **Bookmarks and Navigator:** Setting bookmarks for easy navigation in reports.
- **Buttons and Shapes:** Adding interactive buttons and shapes to enhance user experience.

Data Analysis Expressions (DAX) and Power BI Service

- **DAX Syntax:** Overview of DAX syntax and structure.
- **DAX Operators and Functions:**
 - Key DAX functions for data manipulation: historical, logical, text, collection, and filter functions.
- **Measure Use Case:** Creating and using measures for advanced data analysis.
- **Project Sharing with Power BI Service:**
 - Sharing reports and dashboards via Power BI Service.
 - Generating and sharing dashboard links.
- **Final Project Delivery:** Compiling and presenting a final, interactive project.

Module 3: Database Systems and SQL

Introduction to Databases and SQL

- What is a database?
- Types of databases: Relational vs. NoSQL
- SQL overview: What is SQL?
- Introduction to tables, rows, and columns
- Primary key and foreign key concepts

Data Definition Language (DDL)

- Creating databases and tables
- ALTER, DROP, and TRUNCATE operations
- Data types in SQL (INT, VARCHAR, DATE, etc.)
- Constraints:
 - NOT NULL, UNIQUE, PRIMARY KEY, FOREIGN KEY
 - CHECK, DEFAULT
- Indexes and their importance in database performance

Setting Up Databases and Tables

- Create a simple database and table structure
- Define appropriate data types and constraints
- Use DDL commands to modify database objects

Data Manipulation Language (DML)

- Inserting data into tables (INSERT)
- Updating records in tables (UPDATE)
- Deleting records (DELETE)
- Retrieving data (SELECT)
- Filtering data with WHERE clause

Advanced SELECT Queries

- Ordering data with ORDER BY
- Filtering with LIKE, BETWEEN, IN
- Using aggregate functions:
 - COUNT, SUM, AVG, MIN, MAX
- Grouping data with GROUP BY and filtering groups with HAVING
- Windows functions
 - OVER PARTITION BY
 - Aggregate, Ranking, Value (Lead, Lag), Window Frame

Data Retrieval and Manipulation

- Insert, update, and delete records from tables
- Write complex SELECT queries with filtering, ordering, and aggregation
- Use GROUP BY and HAVING to analyze datasets

Joins and Relationships in SQL

- Types of joins:
 - INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL OUTER JOIN
- Joining multiple tables
- Self joins
- Working with foreign keys
- Combining queries with UNION, INTERSECT, and EXCEPT

Subqueries and Nested Queries

- Introduction to subqueries
- Using subqueries in SELECT, FROM, and WHERE clauses
- Correlated subqueries
- EXISTS vs. IN for subqueries

Working with Joins and Subqueries

- Perform INNER, LEFT, RIGHT, and FULL OUTER joins
- Use subqueries to filter and retrieve data
- Work with self joins and correlated subqueries

Module 4: Python for Data Analysis and Data Wrangling

Why Python and Python Basics

- Why Python?
- How do computers store data?
- Compiled and interpreted programming languages
- Difference between IDE and Jupyter notebooks
- Environments
- Data types and variables
 - INT, STRING, BOOLEAN, FLOAT, DOUBLE
 - Lists, dictionaries, tuples, sets, frozen sets, named tuples
 - String indexing, slicing
 - Print formatting

Conditional Statements and Loops

- IF, ELIF, ELSE
- MATCH CASE
- AND, OR
- Bitwise operators
- While and for loops
- The better alternative of FOR loop

Practical: Python Basics and Control Flow

- Practice with Python basics: variables, data types, and string operations.
- Implement control flow with conditional statements and loops.
- Apply bitwise operations in practical tasks.

Introduction to Object-Oriented Programming (OOP)

- Core concepts: Encapsulation, Abstraction, Inheritance, Polymorphism
- Classes and objects: attributes, methods, constructors
- Encapsulation and access modifiers (public, private, protected)
- Practical example: creating a simple class

Advanced Object-Oriented Programming

- Inheritance: creating subclasses, method overriding
- Polymorphism: method overloading and overriding
- Abstraction: abstract classes and interfaces
- Practical example: building a class hierarchy

Functions and Error Handling

- List comprehension
- Functions and generators
- Iterables, iterators and Decorators
- RETURN and YIELD
- Error handling:
 - Try Except, Raise, Finally

Regular Expressions (Regex)

- RE module
- Basic operations for REGEX:
 - search(), findall(), split(), match(), compile()
- Practical use cases for regex in data analysis

Practical: Functions, Error Handling, and Regular Expressions

- Hands-on tasks with function creation, iterables, and decorators.
- Practice handling errors in Python programs.
- Apply regular expressions to extract and manipulate text data.

Introduction to Web Scraping

- Web scraping with BeautifulSoup and requests
- Basics of HTML, CSS, and JavaScript for web scraping
- Using regex with web scraping
- How to store scraped data in pandas

Advanced Web Scraping Techniques

- Scraping dynamic websites
- Introduction to Selenium:
 - Web drivers
 - Automating web scraping processes
- Advanced scraping techniques with regex
- Interaction with website elements

Practical: Web Scraping

- Implement basic and advanced web scraping techniques using BeautifulSoup and Selenium.
- Practice extracting data from both static and dynamic websites.
- Store scraped data in structured formats (CSV, DataFrames).

Introduction to NumPy

- Overview of NumPy arrays
- Array creation and basic operations
- Array shape, dimensions, and reshaping
- Array indexing and slicing
- Matrix operations in NumPy
- Useful NumPy methods

Advanced NumPy Operations

- Broadcasting in NumPy
- Random number generation
- Sorting, searching, and unique functions
- Linear algebra with NumPy
- Storing and saving NumPy arrays as files

Practical: NumPy Basics and Advanced Operations

- Practice with basic and advanced NumPy operations.
- Implement matrix and linear algebra calculations.
- Store and retrieve NumPy arrays from files.

Introduction to Pandas

- Basic operations with Series and DataFrames
- Reading data from various file types (CSV, Excel, SQL)
- Handling missing data, duplicates, and wrong formats
- Describing and summarizing data with pandas

Data Cleaning and Transformation with Pandas

- String operations in pandas, using regex with pandas
- Removing duplicates and filtering data
- Grouping and aggregation in pandas

Practical: Data Wrangling with pandas

- Load and clean a real-world dataset using pandas.
- Practice string operations, handling missing data, and applying regex.
- Perform grouping and aggregation techniques on datasets.

Advanced Data Manipulation with Pandas

- Reshaping and melting data
- Filtering data with pandas
- Joining and merging DataFrames, Stacking and unstacking
- Using pivot tables for data analysis

Advanced Grouping and Time-Series Data

- Binning and discretization
- Advanced grouping and aggregation techniques
- Working with time-series data in pandas
- Resampling, shifting, and rolling operations

Practical: Advanced Data Analysis with pandas

- Practice with advanced data reshaping, merging, and pivoting techniques.
- Implement time-series analysis on a dataset.
- Explore advanced grouping and aggregation techniques.

Performance and Memory Optimization in pandas

- Techniques for optimizing memory usage in pandas
- Efficiently handling large datasets
- Performance considerations when working with large DataFrames
- Storing and retrieving optimized data formats

Data Visualization and Dashboarding

- Introduction to Data Visualization:
 - Overview of essential principles for effective data visualization.
- Matplotlib and Seaborn:
 - Basic and advanced visualizations (line plots, histograms, box plots, heatmaps). Customization options and styles.
- Interactive Visualizations with Plotly:
 - Building interactive charts and graphs (scatter plots, 3D plots, animated visualizations). Customization options for dynamic and responsive visuals.
- Building Dashboards with Dash:
 - Creating interactive dashboards using Dash.
 - Integrating multiple graphs, charts, and filters for comprehensive data dashboards.
 - Layout and interactivity options for user-driven data exploration.

End-to-End Data Pipeline Project

- Comprehensive Project:
 - Data scraping, cleaning, transformation, and visualization.
- Steps:
 - Scraping data from websites.
 - Cleaning and transforming data with pandas.
 - Creating data visualizations with Matplotlib, Seaborn, Plotly, and Dash.
 - Building an interactive dashboard for data analysis insights.
 - Saving and documenting the project for final delivery.

Statistical Methods and Probability Theory

Descriptive Statistics

- **Introduction:** Overview of descriptive statistics and its role in summarizing data.
- **Mean:** Calculation and interpretation of the mean.
- **Median:** Understanding and calculating the median.
- **Mode:** Finding the mode and understanding its applications.
- **Mean or Median?:** When to use mean vs. median depending on data characteristics.
- **Skewness:** Understanding data skewness.
 - **Practice:** Calculate skewness for sample data.
 - **Solution:** Guided solution to skewness practice.
- **Range and Interquartile Range (IQR):** Measuring data spread.
- **Sample vs. Population:** Distinguishing between sample and population metrics.
- **Variance and Standard Deviation:** Calculating and interpreting variability.
- **Impact of Scaling and Shifting:** How transformations affect measures of spread.
- **Statistical Moments:** Overview of moments (mean, variance, skewness, kurtosis).

Distributions

- **What Is a Distribution?:** Understanding probability distributions.
- **Normal Distribution:** Characteristics and applications.
- **Z-Scores:** Standardizing data and interpreting Z-scores.

Probability Theory

- **Introduction:** Basics of probability and its importance in data analysis.
- **Probability Basics:** Core principles of probability.
- **Calculating Simple Probabilities:**
 - **Practice:** Exercises on calculating basic probabilities.
 - **Quick Solution and Detailed Solution:** Step-by-step answers.
- **Rule of Addition:**
 - **Practice:** Exercises on the addition rule.
 - **Quick Solution and Detailed Solution:** Step-by-step answers.
- **Rule of Multiplication:**
 - **Practice:** Exercises on the multiplication rule.
 - **Solution:** Detailed solution.
- **Bayes' Theorem:** Theory and practical applications.
- **Expected Value:**
 - Concept of expected value and its calculation.
 - **Practice:** Expected value exercises.
 - **Solution:** Guided solution for expected value exercises.
- **Law of Large Numbers:** Explanation and importance.

- **Central Limit Theorem (CLT):**
 - **Theory:** Understanding the CLT.
 - **Intuition and Challenges:** Building intuition for the CLT.
 - **Exercises and Solutions:** Practical applications of the CLT.
- **Binomial Distribution:** Definition, examples, and applications.
- **Poisson Distribution:** Understanding and applying the Poisson distribution.
- **Real-Life Problems:** Applying probability concepts to real-world data.

Hypothesis Testing

- **Introduction:** Importance of hypothesis testing in statistical analysis.
- **What Is a Hypothesis?:** Defining null and alternative hypotheses.
- **Significance Level and P-Value:** Interpreting significance and p-values.
- **Type I and Type II Errors:** Understanding and managing error types.
- **Confidence Intervals and Margin of Error:** Constructing and interpreting intervals.
- **Sample Size and Power Calculations:** Determining appropriate sample size.
- **Performing Hypothesis Tests:**
 - **Practice:** Exercises on hypothesis testing.
 - **Solution:** Guided solution for hypothesis testing exercises.
- **t-Test and t-Distribution:** Theory and application of t-tests.
- **Proportion Testing:** Hypothesis tests for proportions.
- **Important p-z Pairs:** Reference for critical values in hypothesis testing.

Regressions

- **Introduction:** Fundamentals of regression analysis.
- **Linear Regression:** Building and interpreting simple linear models.
- **Correlation Coefficient:**
 - **Practice:** Exercises on calculating and interpreting correlation.
 - **Solution:** Guided solution for correlation exercises.
- **Residual, MSE, and MAE:** Understanding residuals and error metrics.
 - **Practice:** Exercises on MSE and MAE.
 - **Solution:** Guided solution for MSE and MAE practice.
- **Coefficient of Determination (R^2):** Interpreting model fit.
- **Root Mean Square Error (RMSE):**
 - **Practice:** Exercises on RMSE.
 - **Solution:** Solution for RMSE calculations.

Advanced Regression and Machine Learning Algorithms

- **Multiple Linear Regression:** Extending linear regression to multiple predictors.
- **Overfitting:** Understanding overfitting and techniques to mitigate it.
- **Polynomial Regression:** Modeling non-linear relationships.
- **Logistic Regression:** Introduction to classification through logistic regression.
- **Decision Trees:** Basics of decision tree algorithms.

- **Regression Trees:** Tree-based models for regression.
- **Random Forests:** Ensemble method for improved model accuracy.
- **Handling Missing Data:** Techniques for imputation and managing incomplete data.

ANOVA (Analysis of Variance)

- **ANOVA Basics and Assumptions:** Introduction to ANOVA and its assumptions.
- **One-Way ANOVA:** Testing for differences across groups.
- **F-Distribution:** Understanding the F-distribution in ANOVA.
- **Two-Way ANOVA:**
 - **Sum of Squares:** Calculating sum of squares for variance analysis.
 - **F-Ratio and Conclusions:** Interpreting the F-ratio and drawing conclusions.