

Become a Data Analyst

Program Overview



Table of contents

The schedule

Required time commitment	3
Typical weekly schedule	3

The program

Welcome to the Masterschool Data Analysis Program	4
01. Python Fundamentals	4
02. Statistics and Probability Fundamentals	5
03. SQL	7
04. Version Control	8
05. Databases	8
06. Data Preprocessing	9
07. Exploratory Data Analysis	10
08. Data Storytelling	11
09. Analytics with Spreadsheets	12
10. Business Analytics	13
11. Data against intuition: A/B testing	14
12. Complex Analysis	14
13. Airflow and ETL	15
14. Machine Learning	16

The schedule

Required time commitment

Masterschool's courses are designed to be rigorous and intensive. The program requires you to work hard and dedicate *4-6 hours daily, and 25 hours per week*.

Your schedule includes participating in live sessions, reviewing technical materials, and interactive workshops.

Furthermore, our immersive career accelerator requires students to spend additional time on industry projects, networking activities with hiring companies, and other extra-curricular activities.

Typical weekly schedule

This is an example of a typical week in Masterschool. The exact time and days of your weekly sessions will be provided to you during your enrollment.

Week days						Weekend
Live	Study	Live	Live	Live	Study	
Study	Study	Study	Study	Study		
Live						
Live			Live	Live		
5 hrs	4 hrs	4 hrs	5 hrs	5 hrs	2 hrs	-
Min. 25 weekly hours						

 Mandatory lecture

 Mandatory self-study

 Optional sessions

The program

Welcome to the Masterschool Data Analysis Program

We'll take a look at the day-to-day life of a Data Analyst, define the scope of the job and get acquainted with the different analyst roles.

Introduction to Data Analytics

A day in the life of a Data Analyst

Scope

Type of Analyst by Role

01. Python Fundamentals

In this course, you will first learn how to write basic Python statements using building blocks like the print statement, variables, and different data types. You will then look into ways to group and order data types into different data structures.

With an understanding of how data is structured and organized in programming, you will then control the flow of your code with conditionals, loops, and functions.

Then, in the Scripting lesson, you will combine all the concepts you've learned and write and run code in your local environment.

Lastly, you will learn to apply Python to Data Analysis, using some pre-built collections of code, or packages. These include NumPy, a package for efficient scientific computation, and Pandas, a popular Data Analysis package for manipulating data that's built on top of NumPy.

Anaconda and Jupyter Notebooks

Anaconda

Jupyter Notebooks

Data Types and Operators

Data Types

Operators

Built-in Functions

Data Structures

Lists, tuples, sets and dictionaries

Built-in functions

Control Flow

Conditional Statements Boolean Expressions For and While Loops

Break and Continue Zip and Enumerate List Comprehensions

Functions

Functions Docstrings Lambda

NumPy

NumPy Perform operations on ndarrays

Pandas

Series and DataFrames Loading Data NaN values

Iterators and Generator

Iterator vs. list Creating an iterator with a generator

02. Statistics and Probability Fundamentals

Mastering statistics is an essential skill for analyzing data. In this course, you will be learning statistics to summarize data, identify patterns, extract insights and make inference that have a business meaning.

Descriptive Statistics

Evaluate data types and variable types Analyze measures of center

Implement notation Evaluate measures of spread Outliers detection

Probability, Conditional Probability

Calculate the probability of events Conditional Probabilities

Probabilities of multiple dependent events

Distributions

Binomial Distribution

Normal Distribution Theory

Bayes Rule

Describe Bayes Rule

Normalizing

Apply probabilistic inference in a real-world

Python Probability Practice

Probability

Binomial Distributions

Conditional Probability

Bayes Rule

Sampling Distributions and the Central Limit Theorem, Confidence Intervals

Inferential Statistics

Sampling distribution

Notation

Law of Large Numbers

Central Limit Theorem

Bootstrapping

Confidence intervals

Population sampling methods

Population sampling methods

Hypothesys testing and A/B tests

Hypothesis testing

Alternative and Null hypothesis

P-Value

Accept or reject the findings

Large sample sizes and multiple tests

A/B testing

Regression Models

Linear Regression

Multiple Linear Regression

Logistic regression

Confusion matrices

Precision and recall

True positives, false positives

3: SQL

SQL language remains the most efficient way today to extract data. In this course you will learn how to code elegant/optimized SQL queries for carrying out advanced data analysis tasks. SQL is used in virtually every industry so get ready to superpower your skillset.

Basic SQL

DB Types

SQL Query Structure

Joins

Data from Multiple Tables

ERD (primary/foreign Key)

Inner, Outer Joins

Aggregation Functions

Motivation for Aggregating Data

Aggregation Functions

Aggregation by Groups (group By)

Distinct Count Aggregation

Using Conditional Aggregation (case)

Subqueries/ Nested Query and Temporary Tables

Why and When to Use Subquery

Structure and Format of Nested Query

With Statement (temporary Table)

Data Cleaning and Missing Value Handling

Extracting Substrings

Concatenation of Multiple Columns

Handling Null Values

Window Functions

Core Functions

Ranking Functions

Advanced Functions

Advanced Queries Optimization and Tuning

Self Joins, Full Outer Join, Inequality Join

Query Performance Tuning

4: Version Control

Version control is one of the power tools of programming. It allows you to keep track of what you did when, undo any changes you decide you don't want, and collaborate at scale with other people.

Introduction to Git

Working with repositories Git configuration Undo

Working with branches Collaborating Resolving Conflicts

5: Databases

What is Data Base (DB)? How to design it? What differs a well designed DB and a poorly designed one? What should we take into consideration while designing a DB? During this course we'll learn to answer these and many other questions related to DB. Finally, we'll learn how to manage our database and how to improve DB performance.

Database Design

Processing, Storing, and Organizing Data Database Schemas and Normalization

Database Views Database Management

SQL DML: Data Manipulation Language

Inserting Data Updating Data in Postgres Deleting Data in Postgres

Introduction to Data Modeling

Data modeling for relational and NoSQL databases

DM for relational databases with PostgreSQL

DM for NoSQL databases with Apache Cassandra

Relational Data Models

Normalization

Denormalization

Fact/dimension tables

Different schema models

NoSQL Data Models

Basics of NoSQL database design

Denormalization

Primary keys

Clustering columns

The WHERE clause

6: Data Preprocessing

Data preprocessing involves transforming raw data to well-formed data so that advanced analytics techniques can be applied. Raw data is often incomplete and has inconsistent formatting. In this course, you'll be learning all about data wrangling. Data wrangling is the process of gathering your data, assessing its quality and structure, and cleaning it before you do things like analysis, visualization, or build predictive models using machine learning.

Introduction to Data Preprocessing

Motivation of Data Preprocessing

Assessing, Cleaning, Analyzing and Visualizing Data

Data Collection

Collect Data from Flat Files

Scrape Data from Html Files

Use Apis to Collect Data

Process Json Files to Extract Data

Assessing Data

Completeness

Validity

Accuracy

Consistency

Data Cleaning

[Handle Missing Data](#)[Handle Inconsistent Data](#)[Transform Data for Tidiness](#)

Data Preprocessing

[Dealing with Missing Data](#)[Joining Data with pandas](#)[Reshaping Data with pandas](#)

7: Exploratory Data Analysis

Exploratory Data Analysis refers to the process of performing initial investigations on data to discover patterns. In this course you'll learn a diverse set of visualization techniques to unlock preliminary insights and provide you solid leads for further analysis. You'll create, style, and customize a variety of stunning, interactive graphs!

Data Visualization in Data Analysis

[Exploratory vs. Explanatory Analyses](#)[Python Data Visualization Libraries](#)

Design of Visualizations

[Levels of measurement and types of data](#)[Continuous vs. discrete data](#)[Identifying data types](#)[What experts say about visual encodings](#)[Chart Junk](#)[Data-to-ink ratio](#)[Design integrity](#)[Shape, size, and other tools](#)[Using color and designing for color blindness](#)

Univariate Exploration of Data

[Bar charts](#)[Pie charts](#)[Histograms for quantitative variables](#)[Matplotlib and Seaborn packages](#)

Bivariate Exploration of Data

[Heat maps](#)[Scatterplots](#)[Violin plots](#)[Box plots](#)[Clustered bar charts](#)[Faceting](#)[Line plot](#)

Multivariate Exploration of Data

Non-Positional Encodings Color Palettes Faceting in Two Directions
Adaptations of Bivariate Plots Plot Matrices Feature Engineering

Explanatory Visualizations

Steps to tell a story using data Tools and techniques to polish plots
Presenting findings using a slide deck Getting and using feedback

Introduction to Data Visualization with Plotly

Introduction to Plotly Customizing Plots Advanced Customization
Advanced Interactivity

8: Data Storytelling

Knowing how to tell a coherent and convincing story for communicating your analysis results is a key for your success as a data analyst. In this course, you'll learn how to work on your storytelling skills while taking advantage of BI tools. By the end of this course you'll be able to communicate your results in beautiful and self-explanatory dashboards.

Define Problem Statement

Context Problem statement Resolution Limitations and biases

Analysis Roadmap: an issue tree and a ghost deck

Issue Trees Hypothesis Driven Structuring Ghost Deck
Caveats, Limitations, and Bias

Limitations and Biases

Biases in Data Collection Biases in Data Processing Dealing with Missingness

Visualizations and Tying Together

Types of Visualizations

Data Normalization

Eight Data Story Types

Change Over Time

Hierarchy Drill Down

Zoom In / Out

Contrasting Values

Intersections

Different Factors

Outliers

Correlations

Introduction to Tableau

Getting Started with Tableau

Building and Customizing Visualizations

Geographic Data Visualizations

Data changes with time

Trend and forecasting lines

Make your data visually appealing

Dual-axis Graph

Case Study: Analyzing Customer Churn in Tableau

Dashboards and stories

9: Analytics with Spreadsheets

The ability to analyze data rapidly is a powerful skill that helps you make better decisions and iterate faster with your hypotheses. Microsoft Excel/Spreadsheets remains one of the top tools for data analysis due to its rich code-free interface and diverse list of functions. In this course you'll learn explore data from multiple tabs, preprocess your data to a format fit for analysis,

Getting Started

Background on spreadsheets

How to set up a spreadsheet

Major features of spreadsheets

Manipulate Data

Working with cell formulas (text data, math operations, statistical functions)

Performing data operations at the table level (duplicating, splitting, sorting)

Analyze Data

Aggregation functions (like SUM or COUNT)

Named Ranges

Logical Functions (like IF, AND, OR, NOT)

Pivot Tables

Conditional Aggregation functions (such as COUNTIF and SUMIF)

Visualize Data

Pie Charts

Bar Charts

Scatter and Line Plots

Histograms

Box Plots

VLOOKUP

Dashboards

10: Business Analytics

Business analytics techniques can help analysts to measure and optimize important business KPI (Key Performance Indicator). Funnels, a key concept and business analytics, are a great tool for understanding how a sales and marketing operation works. In this course you'll learn to analyze funnels to optimize your KPI, learn how to use cohort analysis for making informed product decisions that reduce churn and increase revenue and finally learn models to evaluate a projected revenue stream from your customers.

Business metrics

Key Performance Indicators

Business Process Flow

Grouping data

Distribution and central tendency

Problem Identification

Identifying Opportunities

Understanding the Market

The Target User

Total Addressable Market

Creating & Evaluating Hypotheses

ROI

Creating a business case

Vision & Strategy

Vision

Strategy

Business Models

Business Models

MVP

Competitive Analysis

Communicational Skills

Active Listening

Storytelling

Being Persuasive

Presentations

Negotiation

Conducting a Retention Cohort Analysis

Types of Retention

Increase engagement

Use cases

Increase frequency

Increase intensity

Engagement state

Retention Analysis

Analyzing Impact of a Churn Rate Across the Business

Lifecycle State

Churn Rate

LTV

11: Data against intuition: A/B testing

Data is the foundation of new waves of productivity growth, innovation and richer customer insights. Companies in every industry are focus on exploiting the data for competitive advantage. Data analysts play a key role in unlocking these in-depth insights, data discoveries and segmentations. In this course, you will learn real-world techniques on customer segmentation, behavioral analytics, business trends running A/B tests and make data-driven decision to help business generate value.

Experimental Design in Python

The Basics of Statistical Hypothesis Testing

Student's t-test

Chi-square test

Pearson correlation

Correction (Bonferroni, Šídák)

Shapiro-Wilk test

ANOVA

Non-parametric tests: Wilcoxon rank-sum test, Spearman correlation

Customer Analytics and A/B Testing in Python

Key Performance Indicators: Measuring Business Success

Analyzing A/B Testing Results

Exploring and Visualizing Customer Behavior

The Design and Application of A/B Testing

Customer Segmentation in Python

Cohort Analysis

Recency, Frequency, Monetary Value analysis

Data pre-processing for clustering

Customer Segmentation with K-means

12: Complex analysis

The knowledge of Complex Analysis methods is essential in order to stand out from the crowd - In this course, you will learn how to apply Graph Theory understanding to find connections and relationships, utilize statistical tools and methods for creating predictions, and have the tools to perform said analysis in every business situation.

Introduction to Network Analysis in Python

Introduction to networks

Important nodes

Structures within network data

Time Series Analysis in Python

Correlation and Autocorrelation

Simple Time Series

Autoregressive (AR) Models

Moving Average (MA) and ARMA Models

Comparing Search Interest with Google Trends

Generating a report

Calculate key metrics

Plotting your answers

13: Airflow and ETL

The use of Airflow is an incredible power that will enable you to perform routine base analysis in a seamless manner, run Pythonic scripts and programs on schedule and create continuous, freshly updated conclusions.

Introduction to Airflow in Python

Running a task in Airflow

Working with DAGs and the Airflow shell

Airflow web interface

Navigating the Airflow UI

Implementing Airflow DAGs

Maintaining and monitoring Airflow workflows

Building production pipelines in Airflow

ETL

Create the ETL foundations

Moving transformed data to a clean table

Insert and delete statements on SQLAlchemy

Load process

14: Machine Learning

Would you be interested in predicting future outcomes based on your data?

This course will introduce you to several fundamental concepts by explaining ML models and algorithms. Next, we will examine the basics of model validation, discuss various validation techniques, and begin to develop tools for creating validated and high-performing models. After the validation, we will build some powerful ARIMA class models to forecast the future. Ultimately, we will work with real datasets and apply techniques in the field of Natural Language Processing to prepare text, words, and documents for further processing. Gain skills in this hugely in-demand and influential field.

Intro To Machine Learning

What is Machine Learning

Machine Learning Models

Deep learning use-cases

Machine Learning for Business

Machine learning and data use cases

Machine learning types

Business requirements and model design

Model validation

Basic Modeling in scikit-learn

Validation Basics

Cross Validation

Selecting the best model with Hyperparameter tuning

Stemming and Lemmatization

Stemming with Python nltk package

Lemmatization with Python nltk package

Applications of Stemming and Lemmatization

ARIMA models

ARMA Models

Statsmodels package to fit ARMA, ARIMA and ARMAX models

Identify promising model orders

Framework for structuring time series projects

Seasonal ARIMA Models

Supervised learning with Scikit learn

Classification

Fine-tuning the model

Preprocessing and pipelines

Scikit-learn metrics that allow to assess model's performance

Regression

Natural language processing fundamentals in Python

Regular expressions & word tokenization

Building a "fake news" detector

Simple topic identification using NLP models

Named-entity recognition on English and non-English text

Machine Learning with tree based models in Python

Classification and Regression Trees

The Bias-Variance Tradeoff

Boosting methods of AdaBoost and Gradient Boosting

Tuning the hyperparameters of a tree-based model

Bagging and Random Forests

“

*Keep exploring.
Keep dreaming.
Keep asking why.
Don't settle for what you already know.
Never stop believing in the power of your
ideas, your imagination,
and your hard work to change the world.*

Barack Obama