Data mentoring track

In this mentoring track we will focus on enhancing our data analytical, data science and machine learning toolkits. An emphasis is laid to use open source tools as far as possible with Python as the preferred language.









Data Analysis basics → Essential Python for Data Analysis → Python for Machine Learning → Apply learning in real life

	Effort required	Topic covered
☐	56 hours	**A. The basics – road to become a data analyst** **This track covers the bare bone basics of being able to analyze data with three powerful and popular tools – Excel, SQL, and Power BI**
☐	8 hours	An introduction to data analysis with Excel – https://www.edx.org/course/introduction-to-data-analysis-using-excel-2
☐	12 hours	A more advanced excel course looking at analyzing and visualizing data https://www.edx.org/course/analyzing-and-visualizing-data-with-excel-2
☐	24 hours	Understand the bread and butter of most organizational data- relational databases and how to query data from them with SQL https://www.edx.org/course/querying-data-with-transact-sql-2
☐	12 hours	Learn about business intelligence and the power of data visualization with the global favorite – Power BI https://www.edx.org/course/analyzing-and-visualizing-data-with-power-bi-2

☐	58 hours	**B. Essential Python for data analysis and data science** **In this track, we will move on and use Python to carry out our analytics work. Python is probably the most powerful and widely used language for data science.**
☐	15 hours	In this beginner course, you will be introduced to the three most popular Python libraries- Numpy, Pandas and Matplotlib https://learn.datacamp.com/skill-tracks/python-programming
☐	17 hours	Wherever your interest lies in Python, you will spend 90% of your time in Pandas – Python's tabular data library. https://learn.datacamp.com/skill-tracks/importing-cleaning-data-with-python
☐	16 hours	We will carry on learning more advanced transformations with Pandas https://learn.datacamp.com/skill-tracks/data-manipulation-with-python
☐	10 hours	Extras to work on during this track Get familiar with Jupyter notebooks, the IDE of choice for data scientists Get familiar with GitHub and version controlling your notebooks
☐	65 hours	**C. Python for Machine learning** **In this track, we will move on the exciting world of machine learning with Python. We will be using Python's most popular ML package – Scikit learn and focus on supervised machine learning**
☐	20 hours	Machine learning has its principles in statistics and hence we will begin by learning the fundamentals of statistical modelling with Python. We will use the NumPy package. https://learn.datacamp.com/skill-tracks/statistics-fundamentals-with-python https://learn.datacamp.com/skill-tracks/machine-learning-fundamentals-with-python
☐	20 hours	We will start our journey into the world of scikit-learn and learn about common linear algorithms https://learn.datacamp.com/skill-tracks/machine-learning-fundamentals-with-python
☐	5 hours	Tree based models can handle non-linear models and are extremely popular. In this module, we will learn to use tree -based models in scikit learn https://learn.datacamp.com/courses/machine-learning-with-tree-based-models-in-python
☐	5 hours	Deep dive into XG-Boost, one of the most powerful algorithms out there https://learn.datacamp.com/courses/extreme-gradient-boosting-with-xgboost
☐	15 hours	**Extras to work on during this track –** Get started on Kaggle ML competitions -the following two are recommended – the legendary titanic dataset for classification and house prices for regression https://www.kaggle.com/c/titanic https://www.kaggle.com/c/house-prices-advanced-regression-techniques