**INTERNSHIP REPORT 2020-21**

**Name Of The Course :-**

1. Programming for Everybody (Getting Started with Python)
2. Data Science with Scala
3. Python for Data Science
4. Introduction to Data Science
5. SQL and Relational Databases
6. Machine Learning with Python

**Student Details :-**

* Name :- Ankit Senjaliya
* Enrollment No. :- 19BT04046
* Program :- Bachelors of Technology
* Branch :- Computer Science And Engineering
* Semester :- III

**Faculty Mentor Details :-**

* Ms. Rujul Desai

**Content Index :-**

* Same as Mentioned below in the “Syllabus” section.

**Name Of Online Platform :-**

1. Coursera
2. Cognitive Class

**About The Course :-**

1. Programming for Everybody (Getting Started with Python) :-

* This course aims to teach everyone the basics of programming computers using Python. We cover the basics of how one constructs a program from a series of simple instructions in Python. The course has no pre-requisites and avoids all but the simplest mathematics. Anyone with moderate computer experience should be able to master the materials in this course. This course will cover Chapters 1-5 of the textbook “Python for Everybody”. Once a student completes this course, they will be ready to take more advanced programming courses. This course covers Python 3.

1. Data Science with Scala :-

* **In this course you will learn about**Basic statistics and data types, preparing data, Feature engineering, Fitting a model and Pipelines and grid search.  Apache Spark™ is a fast and general engine for large-scale data processing, with built-in modules for streaming, machine learning and graph processing. This course shows you how to use Spark’s machine learning pipelines to fit models and search for optimal hyperparameters using a Spark cluster.

1. Python for Data Science :-

* This introduction to Python will kickstart your learning of **Python for data science**, as well as programming in general. This beginner-friendly Python course will take you from zero to programming in Python in a matter of hours.
* Upon its completion, you'll be able to write your own Python scripts and perform basic hands-on data analysis using our Jupiter-based lab environment. If you want to learn Python from scratch, this free course is for you.
* You can start creating your own data science projects and collaborating with other data scientists using [IBM Watson Studio](http://cocl.us/PythonforDataScienceMainPage). When you sign up, you get free access to Watson Studio. Start now and take advantage of this platform.

1. Introduction to Data Science :-

* Find out the truth about what Data Science is. Hear from real practitioners telling real stories about what it means to work in data science. This course was formerly named **Data Science 101**.

1. SQL and Relational Databases :-

* Data is one of the most critical assets of any business. Data needs a database to store and process data quickly. SQL is a language used for a database to query data.
* In this introductory course, you'll learn the basics of the SQL language and the relational databases. You'll start by learning about the relational model and relational model concepts and constraints. By the end of this course, you will have learned and used the five basic SQL statements, some advanced SQL syntax, and join statements.
* This isn't your typical textbook introduction. You're not just learning through lectures. At the end of each module there are assignments, hands-on exercises, review questions, and also a final exam. Successfully completing this course earns you a certificate. So let's get started!

1. Machine Learning with Python :-

* This Machine Learning with Python course dives into the basics of machine learning using an approachable, and well-known, programming language. You'll learn about Supervised vs Unsupervised Learning, look into how Statistical Modelling relates to Machine Learning, and do a comparison of each.
* Look at real-life examples of Machine learning and how it affects society in ways you may not have guessed!
* **Explore many algorithms and models :-**
* Popular algorithms: Classification, Regression, Clustering, and Dimensional Reduction.
* Popular models: Train/Test Split, Root Mean Squared Error, and Random Forests.

**Details of institute offering the course :-**

* All courses are independently published by the platforms or by an instructor/teacher associated with the organizations.

**Name and Details of the instructor :-**

* Charles Russell Severance – Clinical Professor, School Of Information, Expert in Python.

**Syllabus :-**

1. Programming for Everybody (Getting Started with Python) :-

* Week 1 :-

Chapter One - Why we Program?

* Week 2 :-

Installing and Using Python

* Week 3 :-

Chapter One: Why We Program (continued)

* Week 4 :-

Chapter Two: Variables and Expressions

* Week 5 :-

Chapter Three: Conditional Code

* Week 6 :-

Chapter Four: Functions

* Week 7 :-

Chapter Five: Loops and Iteration

1. Data Science with Scala

* Module 1 - Basic Statistics and Data Types
* Vectors and Labelled Points
* Local and Distributed Matrices
* Summary Statistics, Correlations, and Random Data
* Sampling
* Hypothesis Testing
* Module 2 - Preparing Data
* Statistics, Random data and Sampling on Data Frames
* Handling Missing Data and Imputing Values
* Transformers and Estimators
* Data Normalization
* Identifying Outliers
* Module 3 - Feature Engineering
* Feature Vectors
* Categorical Features
* Using Explode, User Defined Functions, and Pivot
* Principal Component Analysis (PCA) in Feature

Engineering

* RFormulas
* Module 4 - Fitting a Model
* Decision Trees
* Random Forests
* Gradient-Boosting Trees
* Linear Methods
* Evaluation
* Module 5 - Pipeline and Grid Search
* Predicting Grant Applications: Introduction
* Predicting Grant Applications: Creating Features
* Predicting Grant Applications: Building a Pipeline
* Predicting Grant Applications: Cross Validation and Model Tuning
* Predicting Grant Applications: Wrapping up

1. Python for Data Science

* Module 1 - Python Basics
* Your first program
* Types
* Expressions and Variables
* String Operations
* Module 2 - Python Data Structures
* Lists and Tuples
* Sets
* Dictionaries
* Module 3 - Python Programming Fundamentals
* Conditions and Branching
* Loops
* Functions
* Objects and Classes
* Module 4 - Working with Data in Python
* Reading files with open
* Writing files with open
* Loading data with Pandas
* Working with and Saving data with Pandas

1. Introduction to Data Science

* Module 1 - Defining Data Science
* What is data science?
* There are many paths to data science
* Any advice for a new data scientist?
* What is the cloud?
* "Data Science: The Sexiest Job in the 21st Century"
* Module 2 - What do data science people do?
* A day in the life of a data science person
* R versus Python?
* Data science tools and technology
* "Regression"
* Module 3 - Data Science in Business
* How should companies get started in data science?
* Tips for recruiting data science people
* "The Final Deliverable"
* Module 4 - Use Cases for Data Science
* Applications for data science
* "The Report Structure"
* Module 5 -Data Science People
* Things data science people say
* "What Makes Someone a Data Scientist?"

1. SQL and Relational Databases 101

* Module 1 - SQL and Relational Databases 101
* Introduction to SQL and Relational Databases
* Information and Data Models
* Types of Relationships
* Mapping Entities to Tables
* Relational Model Concepts
* Module 2 - Relational Model Constraints and Data Objects
* Relational Model Constraints Introduction
* Relational Model Constraints Advanced
* Module 3 - Data Definition Language (DDL) and Data Manipulation Language (DML)
* CREATE TABLE statement
* INSERT statement
* SELECT statement
* UPDATE and DELETE statements
* Module 4 - Advanced SQL
* String Patterns, Ranges, and Sets
* Sorting Result Sets
* Grouping Result Sets
* Module 5 - Working with multiple tables
* Join Overview
* Inner Join
* Left Outer Join
* Right Outer Join
* Full Join

1. Machine Learning with Python

* Module 1 - Supervised vs Unsupervised Learning
* Machine Learning vs Statistical Modelling
* Supervised vs Unsupervised Learning
* Supervised Learning Classification
* Unsupervised Learning
* Module 2 - Supervised Learning I
* K-Nearest Neighbors
* Decision Trees
* Random Forests
* Reliability of Random Forests
* Advantages & Disadvantages of Decision Trees

* Module 3 - Supervised Learning II
* Regression Algorithms
* Model Evaluation
* Model Evaluation: Overfitting & Underfitting
* Understanding Different Evaluation Models
* Module 4 - Unsupervised Learning
* K-Means Clustering plus Advantages & Disadvantages
* Hierarchical Clustering plus Advantages & Disadvantages
* Measuring the Distances Between Clusters - Single Linkage Clustering
* Measuring the Distances Between Clusters - Algorithms for Hierarchy Clustering
* Density-Based Clustering
* Module 5 - Dimensionality Reduction & Collaborative Filtering
* Dimensionality Reduction: Feature Extraction & Selection
* Collaborative Filtering & Its Challenges

**Aim of the Course :-**

1. Programming for Everybody (Getting Started with Python)

* To learn the basics of python

1. Introduction to Data Science

* The art of uncovering the insights and trends in data has been around since ancient times. The ancient Egyptians used census data to increase efficiency in tax collection and they accurately predicted the flooding of the Nile river every year. Since then, people working in data science have carved out a unique and distinct field for the work they do. This field is data science. In this course, we will meet some data science practitioners and we will get an overview of what data science is today.

1. SQL and Relational Databases 101

* We live in a data-driven world now, and with all that data, people have to search through it to find insights to help inform strategy, marketing, operations, and a plethora of other categories. There are a ton of business that use large, relational databases, which makes basic understanding of SQL is a great skill for almost everyone.
* This is a course for not just for beginners interested in learning about SQL and databases but also for data analysts, data scientists, data engineers, app developers, and for business professionals looking to leverage small and big data.

1. Machine Learning with Python

* This Machine Learning with Python course dives into the basics of machine learning using an approachable, and well-known, programming language. You'll learn about Supervised vs Unsupervised Learning, look into how Statistical Modeling relates to Machine Learning, and do a comparison of each. Look at real-life examples of Machine learning and how it affects society in ways you may not have guessed!
* Please pay attention that this course has some hands-on labs which requires that you have working knowledge of Python programming language. If you don't feel you have sufficient skill in Python programming, I recommend you take Python for data science or Data Analysis with Python courses.

**Content of the Course :-**

* Same as the syllabus

**Activities done during the Course :-**

* A lot of quizzes and assignments were provided during the course

**Reading resources provided during the course :-**

* M3-R5 python (programming and problem solving through python)

**Other References :-**

* Google.com
* Youtube.com

**Learning Outcomes :-**

1. Programming for Everybody (Getting Started with Python)

* In this course you will learn about:
* Install Python and write your first program
* Describe the basics of the Python programming language
* Use variables to store, retrieve and calculate information
* Utilize core programming tools such as functions and loops

1. Data Science with Scala

* In this course you will learn about:
* Basic statistics and data types
* Preparing data
* Feature engineering
* Fitting a model
* Pipelines and grid search

1. Python for Data Science

* In this course you will learn about:
* What is Python and why is it useful
* The application of Python
* How to define variables
* Sets and conditional statements in Python
* The purpose of having functions in Python
* How to operate on files to read and write data in Python
* How to use pandas, a must have package for anyone attempting data analysis in Python.

1. Introduction to Data Science

* In this course you will learn about:
* Meet people who work in data science
* Explore definitions of data science
* Learn about data science in a business context
* Discover some use cases for data science

1. SQL and Relational Databases 101

* In this course you will learn about:
* The SQL language
* The relational database model
* The relational model
* Relational model concepts and constraints

1. Machine Learning with Python

* In this course you will learn about:
* How Statistical Modeling relates to Machine Learning and do a comparison of each.
* Real-life examples of Machine learning and how it affects society in ways you may not have guessed!
* In the labs: Use Python libraries for Machine Learning, such as scikit-learn.

**Grades Obtained :-**

* Scored 94% in The Test.

**Copy Of The Certificate :-**

<https://drive.google.com/file/d/1o1UQ2BR0H86T5J3b6hYrAOSy2qhrNkD5/view?usp=sharing>

**Overall Experience**

* A well-designed course, especially for the beginners, who are really enthusiastic and passionate about programming. It really helps you to learn from scratch and it does not require any prerequisites. Everything is broken down making it easy to understand. Had zero programming experience before and this class made it easy and fun to get into. Lecture videos were broken into small parts whereby it's easy to understand and digest.
* However, I'm not sure about the value of the videos titled "Office Hours." They all appeared as self-promotional behavior from the professor, as if to say, "hey look how great I am making this video in Paris." They're really nothing more than that. In one video, he corners a guest speaker at a conference, who appears annoyed after a few minutes and gives the professor a polite brush-off: "other people here would like to talk with me too." After all of these videos, you will ask yourself, "okay, why did Severance make me watch that?" "What was the point?" And you'll shrug your shoulders and say, "another waste of time."
* Some of the examples don't seem to make much practical sense, which would be good for motivation, but they serve their purpose to understand syntax and structure. Some have errors, which is surprising given the huge audience that can help to correct them. There are extra materials about hard- and software developers that are not really connected.
* I couldn't help but get hung up on what "zork" had to do with anything (among a few other examples), or why I would ever need to find the largest letter in a phrase like "hello world." In fact, I think my biggest struggle going through this course was fighting thoughts like, "what am I going to use this for?" I would like to see more content related to solving real world problems, even if it's just in the bonus content.
* Coming at this with no experience in coding at all, this course was good. I think the most difficult part of this class was following the videos. The reason is because real world uses weren't given very often. I struggled with motivation nearing the end of this course because the examples in the videos focused on the "what" the code does and not so much the "why." Because of this, the usefulness of some of the discussed functionalities in python are still unknown to me. To remedy that, I would recommend the exercises in the videos reflect what will be done in the assignments, at least to some degree.
* It is a good and systematic course for beginners - although for real beginners some parts may be quite a bit too geeky, the promise that even if you don't know much about computers you will still be able to follow the course doesn't convince me. If you had some experience with programming, go at your own speed with the course materials and you will be able to finish within a couple of hours - I finished the whole seven weeks in a couple of days, not working on it full time and watching all the videos, which is not really necessary.