



# **SNS COLLEGE OF ENGINEERING**



**Kurumbapalayam(Po), Coimbatore – 641 107**

**Accredited by NAAC-UGC with 'A' Grade**

**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

## **Department of Artificial Intelligence and Data Science**

**Course Name – Foundation of Data Science**

**III Year / V Semester**

**Unit 1 – Introduction**

**Topic 1- Data Scientist**





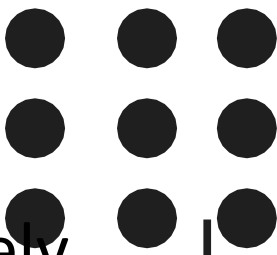
# Data Scientist

## What Does a Data Scientist Do?

- A data scientist analyzes business data to extract meaningful insights. In other words, a data scientist solves business problems through a series of steps, including:
- Before tackling the data collection and analysis, the data scientist determines the problem by asking the right questions and gaining understanding.
- The data scientist then determines the correct set of variables and data sets.
- The data scientist gathers structured and unstructured data from many disparate sources—enterprise data, public data, etc.
- Once the data is collected, the data scientist processes the raw data and converts it into a format suitable for analysis. This involves cleaning and validating the data to guarantee uniformity, completeness, and accuracy.
- After the data has been rendered into a usable form, it's fed into the analytic system—ML algorithm or a statistical model.



# Data Scientist



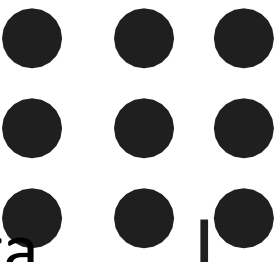
What Does a Data Scientist Do?

This is where the data scientists analyze and identify patterns and trends. When the data has been completely rendered, the data scientist interprets the data to find opportunities and solutions.

The data scientists finish the task by preparing the results and insights to share with the appropriate stakeholders and communicating the results.



# Data Scientist



## Data Science Tools

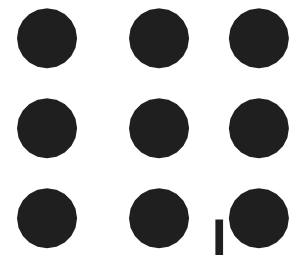
The data science profession is challenging, but fortunately, there are plenty of tools available to help the data scientist succeed at their job.

Data Analysis: SAS, Jupyter, R Studio, MATLAB, Excel, RapidMiner

Data Warehousing: Informatica/ Talend, AWS Redshift

Data Visualization: Jupyter, Tableau, Cognos, RAW, Matlibpro, Power BI, Zoho Analytics

Machine Learning: Spark MLib, Mahout, Azure ML studio, Scikilearn, Pytorch, Tensorflow



**THANK YOU**