

# 10 PROJECTS

to Transition to

# DATA SCIENTIST ROLES





# #1

## Exploratory Data Analysis (EDA) on a Real Dataset

---



### Overview

Choose a dataset of interest and perform in-depth exploratory data analysis, uncovering insights and trends.



### Timeline: 2-3 weeks



### Datasets

- Check Kaggle datasets: <https://www.kaggle.com/datasets>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- Data.gov: <https://www.data.gov/>



### Skills Required

- Data cleaning and preprocessing
- Data visualization with libraries like Matplotlib or Seaborn
- Basic statistical analysis



# #2

## Predictive Modeling with Regression

---



### Overview

Build a predictive model using regression techniques to forecast a numerical target variable.



### Timeline: 3-4 weeks



### Datasets

- Kaggle datasets
- UCI Machine Learning Repository



### Skills Required

- Regression techniques (Linear Regression, Ridge/Lasso Regression, etc.)
- Feature engineering
- Model evaluation and selection



# #3

## Classification with Machine Learning

---



### Overview

Develop a classification model to predict binary or multiclass outcomes.



Timeline: 3-4 weeks



### Datasets

- Kaggle datasets
- UCI Machine Learning Repository



### Skills Required

- Classification algorithms (Logistic Regression, Decision Trees, Random Forest, etc.)
- Feature selection
- Model evaluation metrics (accuracy, precision, recall, F1-score)



# #4

## Time Series Analysis and Forecasting

---



### Overview

Analyze time-dependent data, identify patterns, and create forecasts using time series models.



### Timeline: 4-6 weeks



### Datasets

- Kaggle datasets with time series data
- Financial market data sources
- Weather data sources



### Skills Required

- Time series decomposition
- ARIMA or other time series modeling techniques
- Forecast evaluation and tuning



# #5

## Natural Language Processing (NLP) Project

---



### Overview

Work with text data, performing tasks such as sentiment analysis, text classification, or text generation.



### Timeline: 4-6 weeks



### Datasets

- Kaggle datasets with text data
- Twitter APIs for collecting tweets
- News article APIs or websites



### Skills Required

- Text preprocessing (tokenization, stemming, etc.)
- Feature extraction (TF-IDF, word embeddings)
- NLP libraries (NLTK, spaCy, TensorFlow, etc.)



# #6

## Image Classification with CNNs

---



### Overview

Dive into deep learning by building an image classification model using CNNs.



### Timeline: 6-8 weeks



### Datasets

- CIFAR-10 dataset: <https://www.cs.toronto.edu/~kriz/cifar.html>
- MNIST dataset: <http://yann.lecun.com/exdb/mnist/>
- ImageNet dataset (large and diverse): <http://www.image-net.org/>



### Skills Required

- Convolutional Neural Networks (CNNs) architecture
- Transfer learning (using pre-trained models)
- Image preprocessing and augmentation



# #7

## Recommender System Development

---



### Overview

Create a recommendation engine using collaborative filtering or content-based methods.



Timeline: 4-6 weeks



### Datasets

- MovieLens dataset: <https://grouplens.org/datasets/movielens/>
- Amazon product review datasets
- Last.fm music recommendation dataset: <http://ocelma.net/MusicRecommendationDataset/lastfm-360K.html>



### Skills Required

- Collaborative filtering and content-based methods
- Matrix factorization techniques
- Evaluation metrics for recommendation systems





# #8

## Anomaly Detection in Time Series Data

---



### Overview

Implement anomaly detection algorithms to identify unusual patterns in time series data.



### Timeline: 4-6 weeks



### Datasets

- Numenta Anomaly Benchmark dataset: <https://github.com/numenta/NAB>
- NASA Prognostics Data Repository: <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set>



### Skills Required

- Anomaly detection algorithms (Isolation Forest, One-Class SVM, etc.)
- Feature engineering for anomaly detection
- Evaluation of anomaly detection performance



# #9

## A/B Testing and Hypothesis Testing

---



### Overview

Design and analyze A/B tests to evaluate the effectiveness of changes or interventions.



### Timeline: 3-4 weeks



### Datasets

- Simulated datasets
- Online experimentation platforms for A/B testing



### Skills Required

- Experimental design and control groups
- Hypothesis testing concepts (p-values, confidence intervals)
- Statistical significance and power analysis



# #10

## Deep Learning Project (e.g., Generative Adversarial Networks)



### Overview

Engage in advanced deep learning techniques like GANs for image generation.



### Timeline: 6-8 weeks



### Datasets

- MNIST dataset for GANs: <http://yann.lecun.com/exdb/mnist/>
- CelebA dataset for face generation: <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>
- Artwork or style-specific datasets for artistic GANs






### Skills Required

- Deep learning concepts (Neural Networks, GANs)
- Frameworks like TensorFlow or PyTorch
- Training deep learning models and tuning hyperparameters



# WHY BOSSCODER?

-  **750+** Alumni placed at Top Product-based companies.
-  More than **136% hike** for every **2 out of 3** working professional.
-  Average package of **24LPA.**

The syllabus is most **up-to-date** and the list of problems provided covers all important topics.

Lavanya  
 Meta



Course is very well structured and streamlined to crack any MAANG company

Rahul  




[EXPLORE MORE](#)