



# Data Analytics

with Oasis Infobyte



Project Proposal

# Vision

Embark on a transformative journey in data analytics with our intensive 4-week internship. Designed to equip participants with robust Python programming skills and hands-on experience in real-world data projects, this program is a stepping stone towards a successful career in data analytics.

## Program Highlights:

### 1. Hands-On Data Analytics Projects:

- Our internship is project-centric, providing participants with practical experience by working on real Analytics projects, enhancing their coding proficiency.

### 2. Open-Source Contributions:

- Collaborate with experienced developers on open-source projects

### 3. Resume Enhancement:

- Throughout the program, you'll develop a collection of projects and contributions that will make your resume stand out to potential employers.

### 4. Networking Opportunities:

- Connect with fellow interns, mentors, and industry professionals. Building a strong network can open doors to future career opportunities

### 5. Gradual Skill Progression:

- The program is designed with a gradual learning curve, ensuring that you build upon your knowledge and skills day by day.

### 6. Certificate of Completion:

- Upon successfully completing the program, you'll receive a certificate recognizing your dedication and achievements, a valuable addition to your professional portfolio.

**Note: To successfully complete this internship program, it is essential to accomplish at least three projects from Level 1/Level 2.**

**To be eligible for LOR, you must successfully finish maximum projects from both levels., you must successfully finish maximum projects from both levels.**

<b>Step 1</b>	Review Project Details
<b>Step 2</b>	Commence the Project Development
<b>Step 3</b>	Deploy/Push on Github
<b>Step 4</b>	Create a Video Demonstration of Project Functionality
<b>Step 5</b>	Share the Video on LinkedIn using hashtags #oasisinfobyte, #oasisinfobytefamily, #internship, #python
<b>Step 6</b>	Submit your project carefully in the appropriate batch submission form.
<b>Step 7</b>	Please be patient and await the evaluation of your project; upon completion, you will receive a certificate.

# Table of Contents

1	Project Title: EDA on Retail Sales Data
2	Project Title: Customer Segmentation
3	Project Title: Sentiment Analysis
4	Project Idea: Cleaning Data
5	Project Title: Predicting House Prices with Linear Regression
6	Project Title: Wine Quality Prediction
7	Project Title: Fraud Detection
8	Project Title: Unveiling the Android App Market:
9	Project Title: Autocomplete and Autocorrect Data Analytics
10	Support

# Data Analytics



PROJECT 1 PROPOSAL  
LEVEL 1

## Idea: Exploratory Data Analysis (EDA) on Retail Sales Data

### Description:

In this project, you will work with a dataset containing information about retail sales. The goal is to perform exploratory data analysis (EDA) to uncover patterns, trends, and insights that can help the retail business make informed decisions.

**Dataset 1** [Link](#)

**Dataset 2** [Link](#)

### Key Concepts and Challenges:

1. Data Loading and Cleaning: Load the retail sales dataset.
2. Descriptive Statistics: Calculate basic statistics (mean, median, mode, standard deviation).
3. Time Series Analysis: Analyze sales trends over time using time series techniques.
4. Customer and Product Analysis: Analyze customer demographics and purchasing behavior.
5. Visualization: Present insights through bar charts, line plots, and heatmaps.
6. Recommendations: Provide actionable recommendations based on the EDA.

### Learning Objectives:

- Gain hands-on experience in data cleaning and exploratory data analysis.
- Develop skills in interpreting descriptive statistics and time series analysis.
- Learn to use data visualization for effective communication of insights.

# Data Analytics



PROJECT 2 PROPOSAL  
LEVEL-1

## Idea: Customer Segmentation Analysis

### Project Description:

The aim of this data analytics project is to perform customer segmentation analysis for an e-commerce company. By analyzing customer behavior and purchase patterns, the goal is to group customers into distinct segments. This segmentation can inform targeted marketing strategies, improve customer satisfaction, and enhance overall business strategies.

### Dataset [Link](#)

#### Key Concepts and Challenges:

1. Data Collection: Obtain a dataset containing customer information, purchase history, and relevant data.
2. Data Exploration and Cleaning: Explore the dataset, understand its structure, and handle any missing or inconsistent data.
3. Descriptive Statistics: Calculate key metrics such as average purchase value, frequency of purchases, etc.
4. Customer Segmentation: Utilize clustering algorithms (e.g., K-means) to segment customers based on behavior and purchase patterns.
5. Visualization: Create visualizations (e.g., scatter plots, bar charts) to illustrate customer segments.
6. Insights and Recommendations: Analyze characteristics of each segment and provide insights.

#### Learning Objectives:

- Practical experience with clustering algorithms.
- Data cleaning and exploration skills.
- Visualization techniques for conveying insights.

# Data Analytics



PROJECT 3 PROPOSAL  
LEVEL-1

## Idea: Cleaning Data

### Description:

Data cleaning is the process of fixing or removing incorrect, corrupted, duplicate, or incomplete data within a dataset. Messy data leads to unreliable outcomes. Cleaning data is an essential part of data analysis, and demonstrating your data cleaning skills is key to landing a job. Here are some projects to test out your data cleaning skills:

**Dataset 1** [Link](#)

**Dataset 2** [Link](#)

### Key Concepts and Challenges:

1. Data Integrity: Ensuring the accuracy, consistency, and reliability of data throughout the cleaning process.
2. Missing Data Handling: Dealing with missing values by either imputing them or making informed decisions on how to handle gaps in the dataset.
3. Duplicate Removal: Identifying and eliminating duplicate records to maintain data uniqueness.
4. Standardization: Consistent formatting and units across the dataset for accurate analysis.
5. Outlier Detection: Identifying and addressing outliers that may skew analysis or model performance.

# Data Analytics



PROJECT 4 PROPOSAL  
LEVEL-1

## Idea: Sentiment Analysis

### Description:

The primary goal is to develop a sentiment analysis model that can accurately classify the sentiment of text data, providing valuable insights into public opinion, customer feedback, and social media trends.

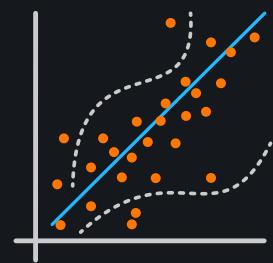
Dataset 1 [Link](#)

Dataset 2 [Link](#)

### Key Concepts and Challenges:

1. Sentiment Analysis: Analyzing text data to determine the emotional tone, whether positive, negative, or neutral.
2. Natural Language Processing (NLP): Utilizing algorithms and models to understand and process human language.
3. Machine Learning Algorithms: Implementing models for sentiment classification, such as Support Vector Machines, Naive Bayes, or deep learning architectures.
4. Feature Engineering: Identifying and extracting relevant features from text data to enhance model performance.
5. Data Visualization: Presenting sentiment analysis results through effective visualizations for clear interpretation.

# Data Analytics



PROJECT 1 PROPOSAL  
LEVEL-2

## Idea: Predicting House Prices with Linear Regression

Dataset [Link](#)

### Description:

The objective of this project is to build a predictive model using linear regression to estimate a numerical outcome based on a dataset with relevant features. Linear regression is a fundamental machine learning algorithm, and this project provides hands-on experience in developing, evaluating, and interpreting a predictive model.

### Key Concepts and Challenges:

1. Data Collection: Obtain a dataset with numerical features and a target variable for prediction.
2. Data Exploration and Cleaning: Explore the dataset to understand its structure, handle missing values, and ensure data quality.
3. Feature Selection: Identify relevant features that may contribute to the predictive model.
4. Model Training: Implement linear regression using a machine learning library (e.g., Scikit-Learn).
5. Model Evaluation: Evaluate the model's performance on a separate test dataset using metrics such as Mean Squared Error or R-squared.
6. Visualization: Create visualizations to illustrate the relationship between the predicted and actual values.

### Learning Objectives:

- Understanding of linear regression concepts.
- Practical experience in implementing a predictive model.
- Model evaluation and interpretation skills

# Data Analytics



PROJECT 2 PROPOSAL  
LEVEL-2

## Idea: Wine Quality Prediction

### Description:

The focus is on predicting the quality of wine based on its chemical characteristics, offering a real-world application of machine learning in the context of viticulture. The dataset encompasses diverse chemical attributes, including density and acidity, which serve as the features for three distinct classifier models.

### Dataset 1 [Link](#)

### Key Concepts and Challenges:

1. Classifier Models: Utilizing Random Forest, Stochastic Gradient Descent, and Support Vector Classifier (SVC) for wine quality prediction.
2. Chemical Qualities: Analyzing features like density and acidity as predictors for wine quality.
3. Data Analysis Libraries: Employing Pandas for data manipulation and Numpy for array operations.
4. Data Visualization: Using Seaborn and Matplotlib for visualizing patterns and insights in the dataset.

# Data Analytics



PROJECT 3 PROPOSAL  
LEVEL-2

## Idea: Fraud Detection

### Description:

Fraud detection involves identifying and preventing deceptive activities within financial transactions or systems. Leveraging advanced analytics and machine learning techniques, fraud detection systems aim to distinguish between legitimate and fraudulent behavior. Key components include anomaly detection, pattern recognition, and real-time monitoring.

### Dataset 1 [Link](#)

### Key Concepts and Challenges:

1. Anomaly Detection: Identifying unusual patterns or deviations from normal behavior within data.
2. Machine Learning Models: Employing algorithms like Logistic Regression, Decision Trees, or Neural Networks for predictive analysis.
3. Feature Engineering: Selecting and transforming relevant features to enhance fraud detection accuracy.
4. Real-time Monitoring: Implementing systems that can detect and respond to fraudulent activities in real-time.
5. Scalability: Designing fraud detection systems capable of handling large volumes of transactions efficiently.

# Data Analytics



PROJECT 4 PROPOSAL  
LEVEL-2

## Idea: Unveiling the Android App Market: Analyzing Google Play Store Data

**Dataset 1** [Link](#)

### Description:

Clean, categorize, and visualize Google Play Store data to understand app market dynamics. Gain in-depth insights into the Android app market by leveraging data analytics, visualization, and enhanced interpretation skills.

#### 1. Data Preparation:

- Clean and correct data types for accuracy.

#### 2. Category Exploration:

- Investigate app distribution across categories.

#### 3. Metrics Analysis:

- Examine app ratings, size, popularity, and pricing trends.

#### 4. Sentiment Analysis:

- Assess user sentiments through reviews.

#### 5. Interactive Visualization:

- Utilize code for compelling visualizations.

#### 6. Skill Enhancement:

- Integrate insights from the "Understanding Data Visualization" course.

-

# Data Analytics



PROJECT 5 PROPOSAL  
LEVEL-2

## Idea: Autocomplete and Autocorrect Data Analytics

### Description:

Explore the efficiency and accuracy of autocomplete and autocorrect algorithms in natural language processing (NLP) through this data analytics project. The objective is to enhance user experience and text prediction by analyzing large datasets and implementing or optimizing autocomplete and autocorrect functionalities.

### Dataset 1 [Link](#)

### Key Concepts and Challenges:

1. Dataset Collection: Gather diverse text data.
2. NLP Preprocessing: Clean and prepare data for analysis.
3. Autocomplete: Implement algorithms for word/phrase predictions.
4. Autocorrect: Optimize algorithms for spelling error correction.
5. Metrics: Define and measure performance metrics.
6. User Experience: Assess impact through feedback and surveys.
7. Algorithm Comparison: Evaluate different models for efficiency and accuracy.
8. Visualization: Use tools for data visualization.



OASIS  
INFOBYTE

# Contact Us for Inquiries

[www.oasisinfobyte.com](http://www.oasisinfobyte.com)

Connect to admin -

<https://taplink.cc/oasisinfobyte.com>

