

Codveda Data Analytics Internship Project Report

Name: Bhavesh Kumar Vishwakamra

1. Introduction

This report presents the work completed during the Data Analytics Internship at Codveda Technologies. The internship focused on applying data analysis, machine learning, and visualization techniques on real-world datasets to extract meaningful insights and build decision-support dashboards.

2. Tools & Technologies

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- Statsmodels
- TextBlob, NLTK
- Power BI

3. Dataset Overview

Dataset Name	Description
Sentiment dataset.csv	Text data for sentiment analysis
Stock Prices Data Set.csv	Historical stock price data
churn-bigml-20.csv	Customer churn dataset (small sample)
churn-bigml-80.csv	Customer churn dataset (large sample)
house_Prediction_Data_Set.csv	House price prediction data
iris.csv	Classic dataset for classification

4. Tasks Completed

- Data Cleaning and Preprocessing
- Exploratory Data Analysis (EDA)
- Regression Analysis
- Time Series Analysis
- K-Means Clustering
- Customer Churn Prediction
- Sentiment Analysis (NLP)
- Power BI Interactive Dashboard

5. Project Summary

Data Cleaning & Preprocessing: Removed missing values, corrected formats, and prepared datasets for analysis.

Exploratory Data Analysis: Visualized distributions, trends, and correlations using charts and graphs.

Regression Analysis: Built models for predicting values such as house prices and evaluated performance metrics.

Time Series Analysis: Analyzed stock price trends over time using moving averages and line plots.

K-Means Clustering: Grouped data into clusters to identify patterns and customer segments.

Customer Churn Prediction: Used classification techniques to predict customer churn and identify risk factors.

Sentiment Analysis: Applied NLP methods to classify customer sentiments as positive, negative, or neutral.

Power BI Dashboard: Designed an interactive dashboard with multiple visuals, slicers, and drill-through capabilities.

6. Key Findings

Task	Key Insight
EDA	Identified major patterns and correlations in datasets
Regression	Models predicted values with acceptable accuracy
Clustering	Data grouped into meaningful segments
Churn Analysis	Highlighted customers with high churn probability
Sentiment Analysis	Classified opinions into positive, neutral, and negative
Dashboard	Provided an interactive view of all analytics results

7. Conclusion

This internship provided valuable hands-on experience in data analytics, machine learning, and business intelligence. By working with multiple datasets and analytical techniques, I developed strong practical skills in extracting insights and presenting them through interactive dashboards.

8. Future Enhancements

- Develop web-based dashboards using Python frameworks
- Deploy machine learning models using APIs
- Enhance NLP with deep learning techniques

9. Reference

GitHub Repository: Codveda Data Analytics Internship Project