

**A REPORT ON THE INTERSHIP TRAINING AT**

**( CHR ENTERPRISES )**

**SUBMITTED BY DIVYA DARSHINI G (212302849)**

**Dr.MGR JANAKI COLLEGE OF ARTS AND SCIENCE FOR  
WOMEN**

**DEPARTMENT OF COMPUTER APPLICATIONS**

**INTERNSHIP REPORT FOR 2023-2024**

<b>NAME</b>	<b>:DIVYA DARSHINI G</b>
<b>REGISTER NO</b>	<b>:212302849</b>
<b>DEPARTMENT&amp;</b>	
<b>YEAR</b>	<b>: BCA&amp;1<sup>st</sup> year</b>
<b>BATCH</b>	<b>: 2023- 2026</b>
<b>SHIFT</b>	<b>: I</b>
<b>INTERNSHIP</b>	<b>: ANALYSIS DATA CODING</b>
<b>COMPANY</b>	<b>: CHR ENTERPRISES</b>



Cell : 9494643768  
9655513091  
Email : chrmanpowers2021@gmail.com

28-05-2024

**INTERNSHIP CONFIRMATION LETTER**

TO

The Head of the Department,  
Department of BCA (Bachelor of Computer application) 1st year,  
Dr.MGR JANAKI COLLEGE OF ARTS AND SCIENCE FOR WOMENS.

Sir/Madam,

**Sub:** Confirmation for Undertaking Internship in our organization,

Greeting from CHR ENTERPRICES, We are providing Internship training to Ms. DIVYA DARSHINI G BCA (Bachelor of Computer Application) 1st year from Dr. MGR Janaki College of Arts and Science for women in the department of Analysis Data Coding duration from (30.05.2024 to 15.06,2024) after Completing Internship we will issue the Internship certificate.

Thanking you



For (CHR ENTERPRISES),

**For CHR ENTERPRISES**

Authorized Signatory   
**Proprietor**





Cell : 9494643768  
9655513091  
Email : chrmanpowers2021@gmail.com

The HOD  
Sathyabama MGR Maligai.  
No.11 & 13, Durgabai Deshmukh  
Road, Raja Annamalai Puram,  
Chennai – 600028.

Respected Sir / Madam,

We are pleased to inform you that we shall be permitting **Ms. DIVYA DHARSHINI G** the following **First Year BCA** student of your college to undergoing “**Analysis Data Coding**” in our organization from **30<sup>th</sup> May 2024** to **15<sup>th</sup> June 2024**.

With best wishes,

Thanking you,



For (CHR ENTERPRISES),

**For CHR ENTERPRISES**

Authorized Signatory   
**Proprietor**



## **ACKNOWLEDGEMENT**

I wish to thank that Dr.MGR JANAKI college of arts and science for women having given me an opportunity to under do training in the company

hereby wish to thank our correspondent Dr.LATHA RAJENDRAN and principal Dr.R.MANIMEGALAI of Dr.MGR JANAKI college of arts and science for women for giving me opportunity taken up institutional training at way to web.

Last but at least I own a deep sense of gratitude to my parents and my friends for their support and co- operation.

# **COMPANY PROFILE**

## **Company Overview**

FutureTech Innovations is a premier provider of cutting-edge IT solutions, specializing in cloud computing, artificial intelligence, and cybersecurity. Our mission is to drive technological advancements and deliver exceptional value to our clients through innovative and reliable services. We pride ourselves on a collaborative and forward-thinking work culture.

## **Our Mission**

Our mission is to drive technological innovation, providing cutting-edge solutions that enhance efficiency and security for businesses worldwide. We strive to foster a future where technology seamlessly integrates with daily life, making the world smarter and more connected.

## **Our Vision**

Our vision is to harness the power of data analysis to drive informed decision-making and innovation. We aim to transform raw data into actionable insights, empowering our team to achieve excellence and sustainable growth.

<b>S.NO</b>	<b>TABLE OF CONTENT</b>	<b>P.NO</b>
<b>1</b>	ABSTRACT	7
<b>2</b>	INTRODUCTION	8
<b>3</b>	KEY STEPS IN DATA ANALYSIS CODING	8
<b>4</b>	Tools and Technologies	10
<b>5</b>	Applications	11
<b>6</b>	Case Study	13
<b>7</b>	Task 1- Sales Data Analysis	14
<b>8</b>	Task 2- Customer Churn Analysis	16
<b>9</b>	Task 3- Movie Rating Analysis	18
<b>10</b>	CONCLUSION	21

## **ABSTRACT**

Data analysis coding has become an essential skill in today's data-driven world, underpinning critical decisions across various domains such as business, healthcare, and scientific research. This abstract provides an overview of the fundamental concepts, methodologies, and tools involved in analysis data coding.

Data analysis coding refers to using programming languages and statistical techniques to interpret and extract insights from data. The process involves several steps: data collection, data cleaning, exploratory data analysis (EDA), data modeling, and result interpretation. Data collection entails gathering relevant information from diverse sources, while data cleaning addresses inconsistencies, missing values, and errors to ensure data quality.

Exploratory Data Analysis (EDA) uses visual methods like histograms and scatter plots to summarize data characteristics and uncover patterns. Data modeling applies statistical or machine learning algorithms to understand relationships and make predictions, with model evaluation techniques like cross-validation ensuring accuracy and reliability. The final stage is interpreting and reporting the findings through comprehensive reports and visualizations, facilitating data-driven decision-making.

Key tools in data analysis coding include Python and R programming languages, with their extensive libraries for data manipulation (Pandas, dplyr), visualization (Matplotlib, ggplot2), and modeling (Scikit-learn, caret). Integrated Development Environments (IDEs) like Jupyter Notebooks and RStudio enhance productivity by providing interactive coding environments.

# INTRODUCTION

Data analysis coding refers to the process of using programming languages and statistical techniques to interpret and extract insights from data. It involves several steps including data collection, cleaning, exploratory data analysis (EDA), modeling, and interpretation. This multidisciplinary field integrates knowledge from computer science, statistics, and domain-specific expertise to transform raw data into actionable knowledge.

## Key Steps in Data Analysis Coding

### 1. Data Collection

- Gathering relevant data from diverse sources such as databases, APIs, web scraping, surveys, and existing datasets.
- Ensuring the quality and relevance of the collected data is crucial for accurate analysis.

### 2. Data Cleaning

- Addressing missing values, correcting errors, and formatting data.
- Techniques include imputation (replacing missing values), normalization (scaling data), and encoding categorical variables.
- Data cleaning is vital as it directly impacts the accuracy of subsequent analysis steps.



### **3. Exploratory Data Analysis (EDA)**

- Summarizing main characteristics of the data using visual methods like histograms, scatter plots, and box plots.
- EDA helps in understanding the distribution, identifying patterns, and detecting anomalies in the data.
- It guides the choice of further analysis techniques and models.

### **4. Data Modeling**

- Applying statistical or machine learning models to the data.
- Common algorithms include linear regression, decision trees, clustering, and neural networks.
- The choice of model depends on the nature of the problem (predictive, descriptive, or prescriptive).

### **5. Model Evaluation**

- Assessing the performance of models using techniques like cross-validation, confusion matrices, and ROC curves.
- Ensuring the reliability and generalizability of the models to new data.

### **6. Interpretation and Reporting**

- Communicating the findings effectively through comprehensive reports and visualizations.
- Making insights accessible to stakeholders, facilitating data-driven decision making.

## Tools and Technologies

- **Programming Languages**

- Python: Widely used for its simplicity and extensive libraries such as Pandas (data manipulation), NumPy (numerical operations), Matplotlib and Seaborn (data visualization), and Scikit-learn (machine learning).
- R: Known for its strong statistical capabilities and packages like dplyr (data manipulation), ggplot2 (data visualization), and caret (machine learning).

- **Integrated Development Environments (IDEs)**

- Jupyter Notebooks: An interactive environment for Python that supports data cleaning, transformation, and visualization.
- RStudio: An IDE for R that facilitates data analysis through an intuitive interface and powerful

- **Data Visualization Tools**

- **Matplotlib, Seaborn**: Python libraries for creating static, animated, and interactive visualizations.
- **ggplot2**: An R package for creating complex and multi-layered visualizations based on the Grammar of Graphics.

## Applications

- **Business Analytics**

Customer segmentation, sales forecasting, market analysis, and optimization of business.

- **Healthcare**

Disease prediction, patient monitoring, treatment optimization, and personalized medicine.

- **Scientific Research**

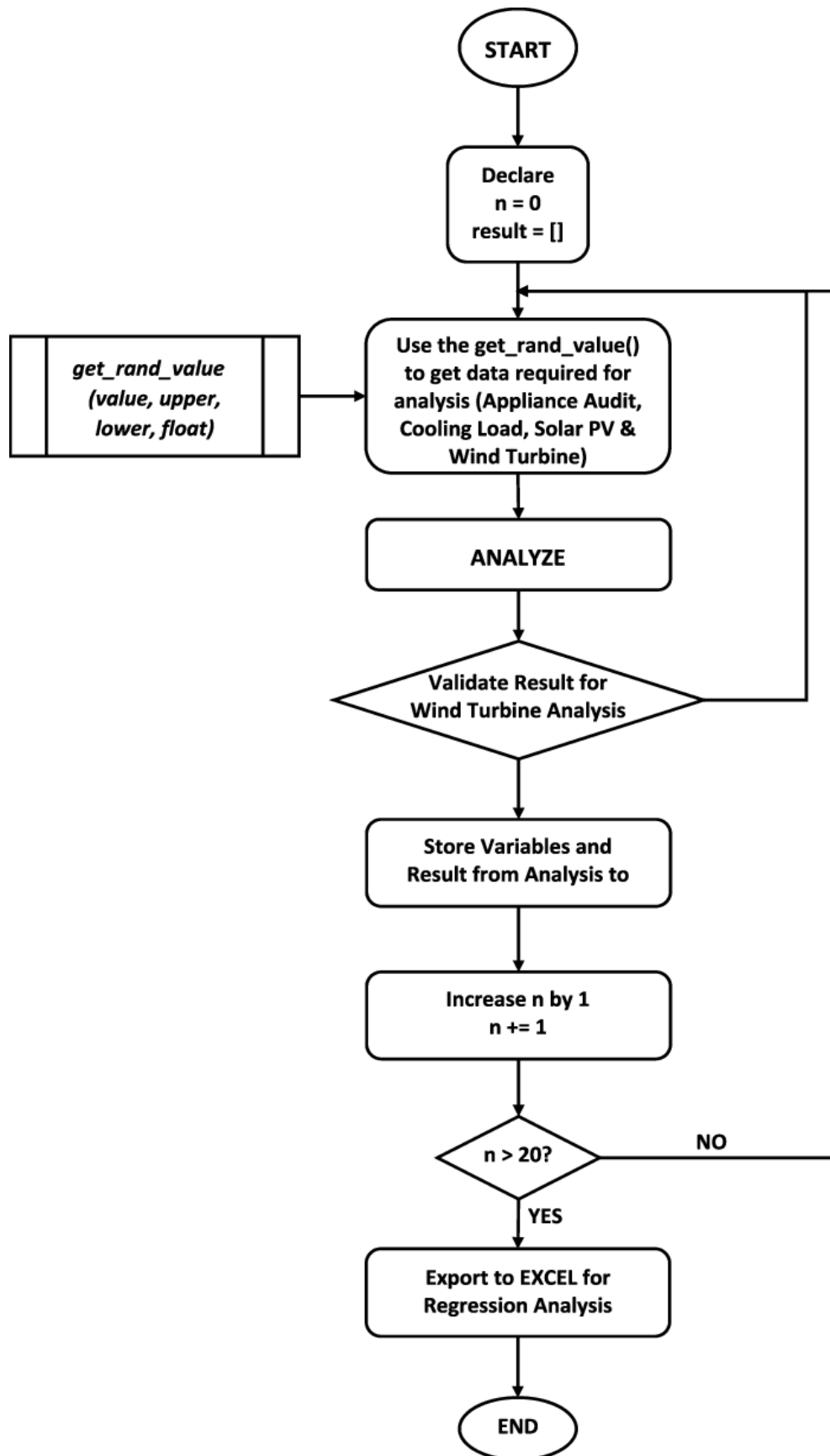
Hypothesis testing, experimental analysis, simulation studies, and data-driven discoveries.

- **Finance**

Risk assessment, fraud detection, investment analysis, and portfolio management.

- **Government and Public Policy**

Policy evaluation, resource allocation, public health monitoring, and socio-economic studies.



## **Case Study: Predicting Customer Churn**

### **1.Objective**

- Predict which customer are likely to churn (leave the service) using historical data

### **2. Data Collection**

- Gather customer data including demographics, usage patterns, service history, and feedback.

### **3. Data Cleaning**

- Handle missing values, correct inconsistencies, and encode categorical variables such as gender and service type.

### **4. Exploratory Data Analysis (EDA)**

- Visualize data distributions and correlations between features and churn.
- Identify key factors influencing customer churn.

### **5. Data Modeling**

- Apply machine learning models such as logistic regression, decision trees, and random forests.
- Use cross-validation to ensure model robustness.

### **6. Model Evaluation**

- Evaluate models using metrics like accuracy, precision, recall, and F1-score.

- Select the best-performing model based on these metrics.

## 7. Interpretation and Reporting

- Present findings through charts and graphs.
- Provide actionable recommendations to reduce churn, such as targeting at-risk customers with special offers.

## Tasks

### Sales Data Analysis

**Objective:** Analyze sales data to determine trends and insights.

**Dataset:** A CSV file containing sales data with columns such as `Date`, `Product`, `SalesAmount`, and `Region`.

### Steps:

1. **Load the Dataset:** Read the CSV file into a Pandas DataFrame.
2. **Data Cleaning:** Handle missing values, if any. Ensure data types are correct (e.g., `Date` should be datetime).
3. **Summary Statistics:** Calculate summary statistics (mean, median, standard deviation) for `SalesAmount`.
4. **Sales Trend:** Plot sales trends over time.
5. **Product Analysis:** Identify the top 5 products by total sales.
6. **Regional Analysis:** Compare sales across different regions.

### Example Code:

### Python

```
import pandas as pd
import matplotlib.pyplot as plt
```

Step 1: Load the Dataset

```
data = pd.read_csv('sales_data.csv')
```

Step 2: Data Cleaning

```
data['Date'] = pd.to_datetime(data['Date'])
data.dropna(inplace=True)
```

Step 3: Summary Statistics

```
print(data['SalesAmount'].describe())
```

Step 4: Sales Trend

```
data.set_index('Date').resample('M')['SalesAmount'].sum().plot()
plt.title('Monthly Sales Trend')
plt.xlabel('Date')
plt.ylabel('Sales Amount')
plt.show()
```

Step 5: Product Analysis

```
top_products =
data.groupby('Product')['SalesAmount'].sum().sort_values(ascending
=False).head(5)
print(top_products)
```

Step 6: Regional Analysis

```
region_sales = data.groupby('Region')['SalesAmount'].sum()
region_sales.plot(kind='bar')
plt.title('Sales by Region')
plt.xlabel('Region')
plt.ylabel('Sales Amount')
plt.show()
```

## **Customer Churn Analysis**

**Objective:** Analyze customer churn data to identify factors affecting churn.

**Dataset:** A CSV file with customer data, including columns such as `CustomerID`, `Churn`, `Tenure`, `MonthlyCharges`, and `TotalCharges`.

### **Steps:**

- 1. Load the Dataset:** Read the CSV file into a Pandas DataFrame.
- 2. Data Cleaning:** Handle missing values and ensure data types are correct.
- 3. Churn Rate:** Calculate the overall churn rate.
- 4. Feature Analysis:** Compare churn vs. non-churn customers for features like `Tenure` and `MonthlyCharges`.
- 5. Correlation Analysis:** Compute correlations between numerical features and churn.
- 6. Visualizations:** Create visualizations to illustrate key findings (e.g., histograms, box plots).



## Example Code:

### Python

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Step 1: Load the Dataset

```
data = pd.read_csv('customer_churn.csv')
```

Step 2: Data Cleaning

```
data['TotalCharges'] = pd.to_numeric(data['TotalCharges'],
errors='coerce')
data.dropna(inplace=True)
```

Step 3: Churn Rate

```
churn_rate = data['Churn'].mean()
print(f'Churn Rate: {churn_rate:.2f}')
```

Step 4: Feature Analysis

```
print(data.groupby('Churn')['Tenure', 'MonthlyCharges'].mean())
```

Step 5: Correlation Analysis

```
correlation = data.corr()
print(correlation['Churn'])
```

Step 6: Visualizations

```
sns.histplot(data, x='Tenure', hue='Churn', multiple='stack')
plt.title('Distribution of Tenure by Churn Status')
plt.show()
```

```
sns.boxplot(x='Churn', y='MonthlyCharges', data=data)
plt.title('Monthly Charges by Churn Status')
plt.show()
```

## **Movie Ratings Analysis**

**Objective:** Analyze movie ratings data to uncover insights about user preferences and movie popularity.

**Dataset:** A CSV file containing movie ratings with columns such as `UserID`, `MovieID`, `Rating`, and `Timestamp`.

### **Steps:**

- 1. Load the Dataset:** Read the CSV file into a Pandas DataFrame.
- 2. Data Cleaning:** Convert `Timestamp` to datetime and handle any missing values.
- 3. Rating Distribution:** Analyze the distribution of ratings.
- 4. Popular Movies:** Identify the top 10 most-rated movies.
- 5. User Behavior:** Analyze the rating behavior of users (e.g., average rating per user).
- 6. Time Analysis:** Examine how ratings have changed over time.

## Example Code:

### Python

```
import pandas as pd
import matplotlib.pyplot as plt
```

Step 1: Load the Dataset

```
data = pd.read_csv('movie_ratings.csv')
```

Step 2: Data Cleaning

```
data['Timestamp'] = pd.to_datetime(data['Timestamp'], unit='s')
data.dropna(inplace=True)
```

Step 3: Rating Distribution

```
rating_counts = data['Rating'].value_counts().sort_index()
rating_counts.plot(kind='bar')
plt.title('Rating Distribution')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.show()
```

Step 4: Popular Movies

```
top_movies = data['MovieID'].value_counts().head(10)
print(top_movies)
```

Step 5: User Behavior

```
user_avg_ratings = data.groupby('UserID')['Rating'].mean()
print(user_avg_ratings.describe())
```

#### Step 6: Time Analysis

```
data.set_index('Timestamp').resample('M')['Rating'].mean().plot()  
plt.title('Average Rating Over Time')  
plt.xlabel('Date')  
plt.ylabel('Average Rating')  
plt.show()
```

**These tasks should provide a solid foundation for analyzing various datasets, helping you develop skills in data cleaning, manipulation, and visualization.**

## **Conclusion**

Data analysis coding is a powerful tool for transforming raw data into meaningful insights. By following a structured approach involving data collection, cleaning, EDA, modeling, and interpretation, valuable information can be extracted to drive informed decision-making. Mastery of data analysis coding, supported by robust tools and technologies, is essential in today's data-centric world, enabling organizations to gain a competitive edge and achieve their objectives efficiently.