Excelerate Internship Team 9 Internship

Welcome to the Excelerate Internship Team 9 journey.

This internship focuses on mastering AI data analysis skills through hands-on projects.

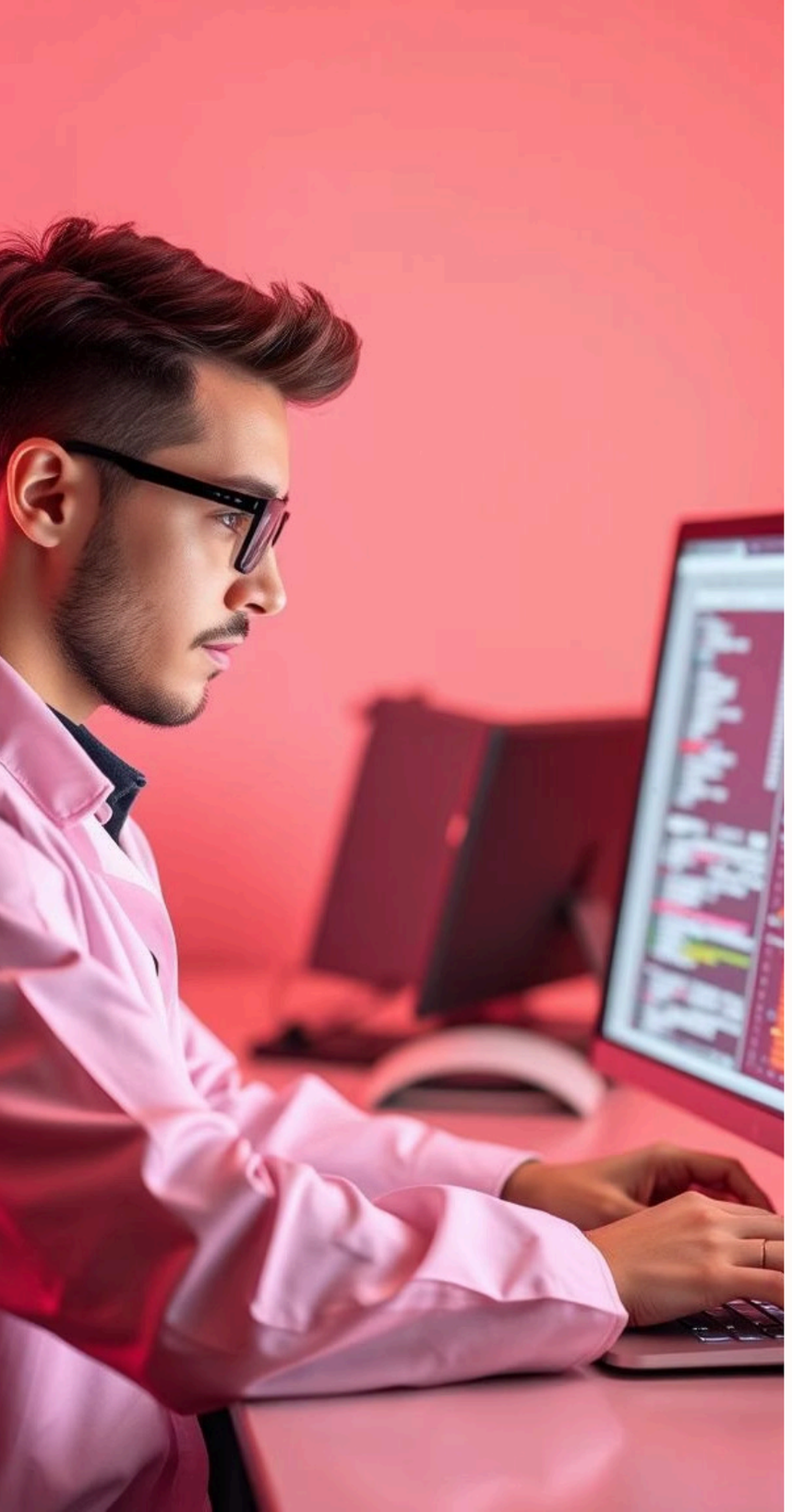
Prepare to explore data preprocessing, analysis, and predictive modeling techniques.

Kiroll Kirollos
Ayan Afridi
Samiat Bhadmus

Understanding the Role and Internship Deliverables of an AI Data Analyst

An AI Data Analyst specializes in analyzing large datasets using artificial intelligence techniques to extract insights, identify trends, and support decision-making. They use machine learning, statistics, and visualization tools to preprocess data, build predictive models, and generate actionable reports.

The internship culminates in a comprehensive report and presentation covering student engagement analysis, predictive modeling, churn analysis, recommendations, and a prototype recommendation system developed over four weeks.



Week 1: Data Understanding and Preprocessing

Understanding the Dataset

Examine dataset structure, identify key variables, and understand data types to build a foundation for preprocessing and feature engineering.

Data Cleaning and Validation

Address missing values, remove duplicates, and correct inconsistencies to ensure dataset accuracy and readiness for analysis.

Feature Engineering

Create new features from cleaned data, such as deriving age or calculating durations, to enhance analytical value.

Report Creation

Compile a detailed report describing dataset characteristics, cleaning steps, and new features to support further analysis.

Week 2: Exploratory Data Analysis (EDA)

Data Cleaning

Ensure data consistency by handling missing values, removing duplicates, and verifying data types.

Exploratory Data Analysis

Analyze signup and completion rates, identify patterns, correlations, and outliers to understand trends.

Data Visualization

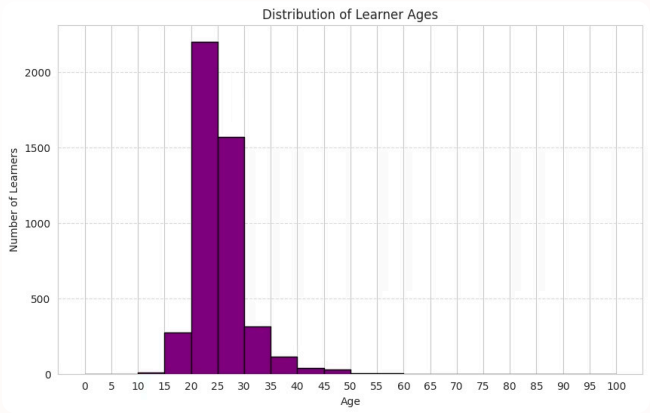
Create multiple visualizations like bar charts, line charts, and scatter plots to communicate data insights clearly.

Insight Extraction

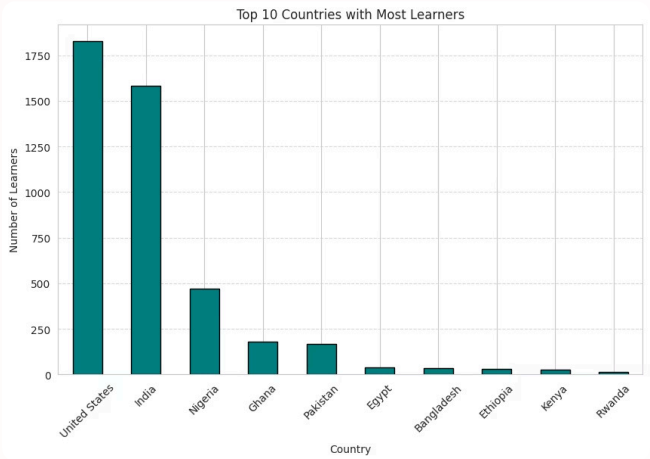
Summarize key findings, discuss anomalies, and provide actionable recommendations to inform decisions.

Visualization Examples

Signup Month



Age Distribution



Top 10 Countries

Week 3: Churn Analysis and Predictive Modeling

Building Predictive Models

Develop and train machine learning models to forecast student drop-offs accurately.

Evaluating Model Performance

Assess models using metrics like accuracy and precision to ensure reliable predictions.

Analyzing Churn Factors

Identify key factors contributing to student drop-offs to understand and address retention challenges.

Report Structure

Include objectives, methods, findings, and recommendations in a professional, well-organized PDF report.

Churn Analysis and Predictive Modeling

```python
# Encode 'Status Description'
le = LabelEncoder()
df['Status_Code_Label'] = le.fit_transform(df['Status Description'])

# Define engagement classes
def classify_engagement(row):
    if row['Engagement Category'] == 'Poor':
        return 2
    elif row['Engagement Score (%)'] < 60 and row['Engagement Score (%)'] >= 40:
        return 1
    else:
        return 0

df['engagement_status'] = df.apply(classify_engagement, axis=1)

# Feature selection
features = ['Status_Code_Label', 'Engagement Score (%)']
X = df[features]
y = df['engagement_status']
```

Churn Factors

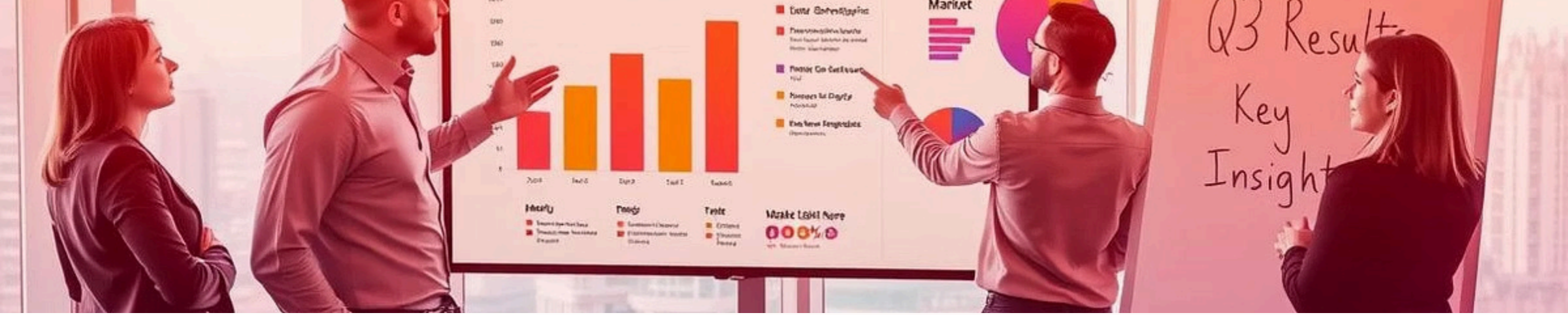
```
Validation Accuracy: 1.0
[[248   0   0]
 [  0 101   0]
 [  0   0 564]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       248
           1       1.00      1.00      1.00       101
           2       1.00      1.00      1.00       564

    accuracy                           1.00       913
   macro avg       1.00      1.00      1.00       913
weighted avg       1.00      1.00      1.00       913

✅ Model and encoders saved.
```

Evaluation Metrics

```python
# Predict
prediction = rf.predict(input_features)[0]
label_map = {0: 'Safe', 1: 'At Risk', 2: 'Dropped Off'}
print(f"📊 Predicted Status: {label_map[prediction]}")
```

```
Enter student's Status Description: Team Allocated
Enter student's Engagement Score (%): 71
📊 Predicted Status: Safe
```

Prediction Model



Week 4: Final Report and Presentation

1

Comprehensive Report

Summarizes all work including engagement analysis, predictive modeling, churn insights, and recommendations.

2

Presentation

Highlights key insights, outcomes, and proposed strategies based on data analysis.

3

Recommendation System Prototype

Demonstrates a functional model to improve student engagement using data-driven suggestions.

Learning Outcomes: Data Cleaning and EDA Skills

Data Cleaning Proficiency

Enhance skills to identify and resolve data quality issues for reliable analysis.

Analytical Thinking

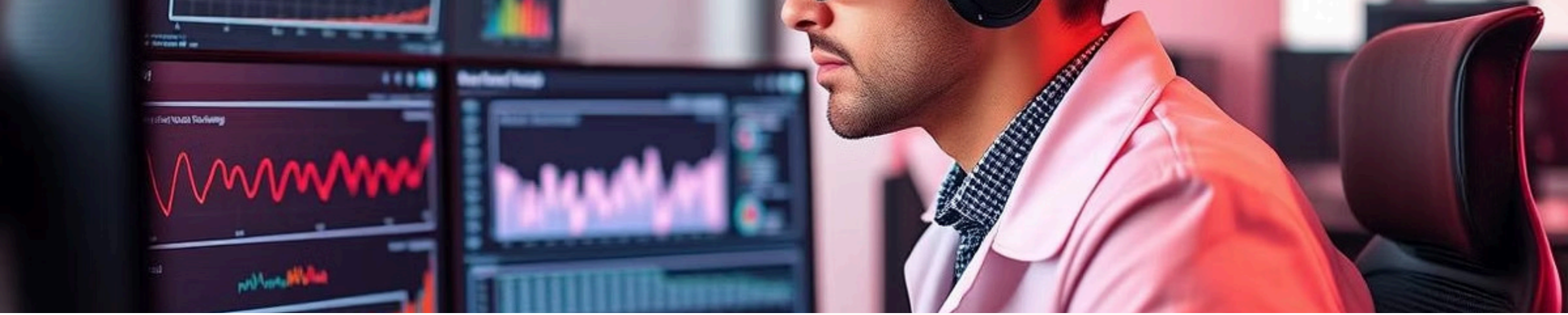
Develop ability to uncover trends, patterns, and outliers for deeper insights.

Effective Communication

Improve report writing and visualization skills to clearly convey findings.

Foundational EDA Skills

Build a solid base in exploratory data analysis techniques for advanced tasks.



Learning Outcomes: Predictive Modeling and Churn Analysis

1 Predictive Modeling Skills

Build and evaluate machine learning models to forecast outcomes effectively.

2 Churn Analysis Expertise

Gain insights into factors influencing student drop-offs and retention strategies.

3 Effective Report Writing

Compile complex analyses into clear, structured reports for stakeholders.

4 Actionable Insights

Translate data findings into practical recommendations to improve engagement.

Conclusion

Key Takeaways

Internship develops skills in data cleaning, EDA, predictive modeling, and churn analysis.

Impactful Deliverables

Final report and presentation translate insights into strategic retention solutions.

Next Steps

Submit final work on time and leverage feedback for future data projects.