

# Presentation Script

## Student Performance EDA

Flashcard Guide for Oral Presentation



Bachir




Mouhamed

### How to Read This Script

- **BLUE boxes** = Bachir's parts
- **GREEN boxes** = Mouhamed's parts
- **Red bold text** = Key words to emphasize
- 0Blue numbers = Important statistics
- 👉 *[PAUSE]* = Take a short pause
- 🖐 *[Point]* = Hand gesture instruction

### BACHIR'S PART

#### Slide 1: Title (30 sec)

 **Goal:** Introduce yourself and the project

"Good morning everyone." 👉 *[PAUSE]*

"My name is **Bachir Zekhnine**, and this is my colleague **Mouhamed Souici**."

👉 *[PAUSE]*

"Today, we will present our **Exploratory Data Analysis** project on **Student Performance**."


👉 *[PAUSE]*

"We analyzed what factors influence how well students perform academically."

#### Tip

Smile, make eye contact, speak slowly and clearly!

#### Slide 2: Project Overview (1 min)

 **Goal:** Explain objective, questions, and methodology

“Let’s start with our **project overview**.” 🖐️ *[Point to slide]*

“Our main **objective** was to analyze student performance and understand the factors that influence **academic success**.”

👉 *[PAUSE]*

“We asked **three key questions**:”

- “**First**: Do scores vary by gender?”
- “**Second**: What is the impact of socioeconomic status?”
- “**Third**: Does test preparation make a difference?”

👉 *[PAUSE]*

“Our **methodology** included four steps.” 🖐️ *[Count with fingers]*

1. “Data cleaning”
2. “Univariate analysis”
3. “Bivariate relationships”
4. “And clustering analysis”

### 📄 Slide 3: The Dataset (1 min)

💡 **Goal:** Describe the data you worked with

“Now let me introduce our **dataset**.” 🖐️ *[Point to slide]*

“We used the **Student Performance Dataset**, which contains academic records along with demographic and socioeconomic factors.”

👉 *[PAUSE]*

“The dataset has:” 🖐️ *[Count with fingers]*

- “01,000 student records”
- “08 features”
- “0Zero missing values”

👉 *[PAUSE]*

“The features are divided into:”

“**Categorical** variables: Gender, Race, Parental Education, Lunch type, and Test Preparation status.”

“**Numerical** variables: Math, Reading, and Writing scores.”

### 🔧 Slide 4: Data Cleaning (45 sec)

💡 **Goal:** Explain how you prepared the data

“Before analysis, we **cleaned** our data. Here are the steps we followed:”

👉 *[PAUSE]* 🖐️ *[Count with fingers]*

1. “We checked for **missing values** — there were **none**”

2. “We **removed** any duplicate records”
3. “We created new features through **feature engineering**”

👉 [PAUSE]

“Specifically, we created:”

- “A **total score** by adding math, reading, and writing scores”
- “An **average score** by dividing the total by three”

“After cleaning, we retained all 01,000 records.”

### 📊 Slide 5: Score Distributions (1 min)

💡 **Goal:** Explain the histogram and what it shows

“Let’s look at our **univariate analysis**, starting with score distributions.”

👉 [PAUSE] 🖐️ [Point to histogram]

“As you can see in this **histogram**:”

- “The mean **math** score is approximately 066”
- “The mean **reading** score is about 069”
- “The mean **writing** score is around 068”

👉 [PAUSE]

“Students generally perform **better** in reading and writing compared to math.”

“The distributions are slightly **left-skewed**, meaning more students score above the mean.”

### 👥 Slide 6: Categorical Variables (45 sec)

💡 **Goal:** Show gender and test prep distribution, then transition to Mouhamed

“Now let’s examine the **categorical** variables.” 🖐️ [Point to chart]

👉 [PAUSE]

“For **gender** distribution:”

- “Females make up 051.8% of the dataset”
- “Males make up 048.2%”

👉 [PAUSE]

“For **test preparation**:”

- “Only 035.8% completed the test prep course”
- “064.2% did **not** complete it”

“This is an **important observation** that we’ll explore further.”

👉 [PAUSE]

#### Transition

“Now, I’ll hand over to **Mouhamed** for the bivariate analysis.”



## MOUHAMED'S PART



## Slide 7: Correlations (1 min)

**Goal:** Explain the correlation matrix

"Thank you, Bachir." [PAUSE]

"Now let's look at the **relationships** between variables."

[Point to correlation matrix]

[PAUSE]

"Our **correlation matrix** reveals some interesting findings:"

- "Reading and Writing scores have a very strong correlation of 00.95"
- "Math and Reading have a **moderate** correlation of 00.82"
- "Math and Writing have a **moderate** correlation of 00.80"

[PAUSE]

"This tells us that **verbal skills** — reading and writing — are highly linked."

"If a student is good at reading, they're likely good at writing too."

"However, **math skills** are more independent."



## Slide 8: Test Preparation Impact (1 min)

**Goal:** Show that test prep helps ALL subjects

"Here's a **major finding**:" [PAUSE]

"Test preparation completion improves scores across ALL subjects."

[PAUSE]

"Remember, only 035.8% of students completed test prep."

"These students showed:"

- "**Higher** Math scores"
- "**Higher** Reading scores"
- "**Higher** Writing scores"

[PAUSE]

"The 064.2% who didn't complete test prep showed **lower** performance overall."

"This represents a missed opportunity."

[PAUSE]

"The implication is clear: **Structured support** provides tangible benefits."

"Schools should **expand access** to test preparation programs."



## Slide 9: Parental Education (45 sec)

**Goal:** Show education level ranking

“Another key finding.” 👉 [PAUSE]

“**Higher parental education** correlates with better student performance.”

👉 [PAUSE] 🖐️ [Point to table]

“Look at this ranking.”

1. “Students with parents holding **Master’s degrees** perform the highest”
2. “Followed by **Bachelor’s** degree holders”
3. “Then Associate’s degree, some college, high school”
4. “Students whose parents have only **some high school** education perform the lowest”

👉 [PAUSE]

“This shows the importance of **educational background** in families.”

### 👤 Slide 10: Gender Gap (1 min)

💡 **Goal:** Explain gender differences in subjects

“Now let’s discuss **gender differences** in performance.”

👉 [PAUSE] 🖐️ [Point to chart]

“**Females** show:”

- “0+7.1 points higher in **Reading**”
- “0+9.2 points higher in **Writing**”
- “But 0-5.2 points lower in Math”

👉 [PAUSE]

“**Males** show:”

- “0+5.2 points higher in **Math**”
- “But lower verbal scores”

👉 [PAUSE]

“So gender does influence subject-specific performance, but the overall differences are **modest**.”

### 🏠 Slide 11: Socioeconomic Impact (1 min)

💡 **Goal:** This is the MOST IMPORTANT finding!

“This is perhaps our **most important** finding.” 👉 [PAUSE]

“Lunch type serves as a **proxy** for socioeconomic status.”

👉 [PAUSE] 🖐️ [Point to chart]

“Students with **standard lunch** — meaning higher socioeconomic status — scored:”

- “0+11.1 points higher on **average**”
- “0+12.7 points higher in **Math**”
- “0+11.5 points higher in **Reading**”
- “0+11.8 points higher in **Writing**”

👉 [PAUSE]

"This gap is significant and shows that **economic factors** strongly predict academic performance."

## 🏆 Slide 12: Best Performing Group (45 sec)

💡 **Goal:** Show combined effect of gender + SES

"When we **combine** factors, we find that." 👉 [PAUSE]

"Females with standard lunch are the **highest performers**."

👉 [PAUSE]

"The ranking is:" 🖐️ [Count with fingers]

1. "Female + Standard lunch: **Highest**"
2. "Male + Standard lunch: High"
3. "Female + Free/Reduced lunch: Moderate"
4. "Male + Free/Reduced lunch: **Lowest**"

👉 [PAUSE]

"This shows a **compound effect** of gender and socioeconomic status."

"Male students with free or reduced lunch need the **most support** and intervention."

## 📊 Slide 13: Clustering (1 min)

💡 **Goal:** Explain the 3 student groups

"We also performed **K-Means clustering** to segment students into groups."

👉 [PAUSE] 🖐️ [Point to cluster visualization]

"We identified **3 clusters**:"

1. "**High achievers** — about 015% of students"
2. "**Average performers** — about 070%"
3. "**At-risk students** — about 015%"

👉 [PAUSE]

"Each cluster requires different approaches:"

- "High achievers should get **advanced programs** and gifted courses"
- "Average performers need continued support and **monitoring**"
- "At-risk students require **intervention programs** and tutoring"

👉 [PAUSE]

"This clear separation enables **targeted educational strategies**."

## 📄 Slide 14: Predictors Ranked (45 sec)

💡 **Goal:** Summarize what matters most

"To summarize **what matters most**, here are the factors ranked by impact:"

👉 [PAUSE] 🖐️ [Count with fingers]

1. "**Socioeconomic status** is the strongest predictor"
2. "**Test preparation** has a strong positive effect"
3. "**Parental education** shows clear correlation"
4. "Race and ethnicity show moderate variation"
5. "**Gender** differences are modest and subject-dependent"

👉 [PAUSE]

"The **critical insight**:"

"Economic factors outweigh all others."

"Addressing socioeconomic barriers should be the **top priority**."

## ★ Slide 15: Key Takeaways (45 sec)

💡 **Goal:** Quick summary of everything

"Let me quickly **summarize** what we learned:" 👉 [PAUSE]

- "**Socioeconomic status** is the strongest predictor of success"
- "**Test preparation** improves all scores"
- "Reading and writing are highly linked with a correlation of 00.95"
- "**Parental education** correlates with performance"
- "Gender differences exist but are **modest**"
- "**Clustering** enables targeted interventions"

👉 [PAUSE]

"Our **bottom line**:"

*"Address socioeconomic barriers and expand test prep access for maximum impact."*

## ♥ Slide 16: Thank You (30 sec)

💡 **Goal:** End professionally and invite questions

"**Thank you** for your attention." 👉 [PAUSE]

"We hope this analysis provides useful insights into student performance factors."

👉 [PAUSE]

"If you have any **questions**, we'd be happy to discuss them."

“You can also reach us at the email shown on the slide, or check our **GitHub repository** for the full code.”

👉 *[PAUSE]*

“Thank you!”

😊 Final Tip

Smile and wait for applause! Be ready for questions.

## ? POTENTIAL QUESTIONS

### Q1: Why K-Means with 3 clusters?

**Answer:**

“We used the **elbow method** to determine the optimal number of clusters.”

“Three clusters provided the **best separation** while remaining interpretable.”

### Q2: What tools did you use?

**Answer:**

“We used **Python** with:”

- “**Pandas** for data manipulation”
- “**Matplotlib** and **Seaborn** for visualization”
- “**Scikit-learn** for clustering”

### Q3: What do you recommend to schools?

**Answer:**

“We recommend:”

1. “**Expanding access** to test preparation programs”
2. “Providing **additional support** for students from lower socioeconomic backgrounds”
3. “Using **clustering** to identify at-risk students early”

### Q4: Are there limitations?

**Answer:**

“Yes. This is **observational data**, so we can identify **correlations** but not prove causation.”

“Also, the dataset is from one specific context and may not **generalize** to all populations.”

## 📋 QUICK REFERENCE CARD

### # Key Numbers to Remember

- Records: 01,000
- Features: 08
- Missing values: 00
- Math mean: 066
- Reading mean: 069
- Writing mean: 068
- Females: 051.8%
- Males: 048.2%
- Test prep completed: 035.8%
- Not completed: 064.2%
- Read-Write correlation: 00.95
- Math-Read correlation: 00.82
- SES gap (average): 0+11.1 pts
- High achievers: 015%
- Average: 070%
- At-risk: 015%

### 💡 Presentation Tips

1. **Eye contact** — Look at your audience, not the slides
2. **Pace** — Speak slowly, English is not your first language
3. **Point at visuals** — When showing graphs, point to what you're discussing
4. **Transitions** — Use "Now let's move on to..." or "As you can see..."
5. **Practice** — Rehearse together 2-3 times before the presentation
6. **Breathe** — Take deep breaths to stay calm

Good luck, Bachir & Mouhamed! 👍