

---

## The Mini-Project Student Handout

### Mini-Project: The VidyaVerse Viral Challenge Basics of Data Analytics (BDA) | Mini Project

---

#### Project Title:

"The VidyaVerse Viral Challenge: From Clicks to Course Champions"

---

#### ✓ Objective:

You are a newly onboarded Data Analyst at **VidyaVerse**, a buzzing Ed-Tech startup. The company has just concluded its ambitious "#DataDreamer" social media campaign. The campaign went viral, but now the C-suite wants to know: **"Did our viral buzz translate into valuable students, or was it just vanity metrics?"** Your mission, as part of a special task force, is to:

1. **Lead a data-driven investigation**, starting from raw, siloed datasets.
  2. **Apply a full data analytics workflow**, from data cleaning and integration to final analysis.
  3. **Leverage a modern toolset**, including SQL, Python, Notion AI, and MindMup, to manage, analyze, and visualize your findings.
  4. **Craft and present a compelling data story** to the Head of Growth, making a clear recommendation for the next multi-million-rupee campaign.
- 

#### 📍 Scenario Background:

VidyaVerse's "#DataDreamer" campaign flooded Instagram, YouTube, and LinkedIn. It generated massive clicks and impressions. But the key questions remain unanswered:

- Which platform's viral reach brought in students who actually *finish* the course?
- Was the high ad spend on one platform justified by the quality of students it attracted?
- Can we create a "Success Profile" of a student who is most likely to complete our course?

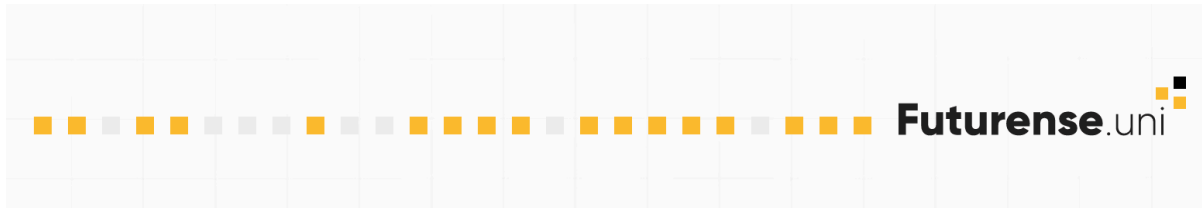
The Head of Growth, Priya Sharma, needs you to cut through the noise and deliver a definitive, data-backed verdict.

### Tools to Be Used:

Tool	Purpose	Use Case in Project
<b>Notion AI</b>	Project Management & Documentation	Create a dynamic project workspace, auto-generate task lists, and summarize key findings in a collaborative doc.
<b>MySQL Workbench</b>	Database Management & SQL Queries	Ingest raw data and perform powerful SQL JOINS to create a unified view of the customer journey.
<b>Google Colab + Python</b>	Data Cleaning & Deep Analysis	Wrangle messy data with Pandas and conduct the core statistical analysis to uncover insights.
<b>MindMup</b>	Brainstorming & Hypothesis Mapping	Visually map out your initial hypotheses and the data points needed to validate them.
<b>ChatGPT / Gemini</b>	Code Assistance & Narrative Crafting	Get help with tricky code snippets and transform your analytical findings into a persuasive story.
<b>Perplexity AI</b>	Industry Benchmarking	Research what defines "good" engagement and ROI in the competitive Ed-Tech landscape.

### Datasets Provided (as CSV files):

-  **campaign\_performance.csv**: Raw export from the ad platforms.



- campaign\_id, ad\_date, platform (e.g., 'Instagram', 'YouTube', 'insta'), ad\_spend, clicks, impressions.
- ☒ **student\_enrollments.csv**: Data from the student information system.
  - student\_id, signup\_date, course\_name, completion\_percentage, source\_campaign\_id.

---

### End Goal:

- **A "Single Source of Truth"**: A clean, merged master dataset.
- **An "Insight Dashboard"**: A clear summary of which platform delivered the highest ROI.
- **A "Strategic Playbook"**: A confident recommendation for future marketing investments.

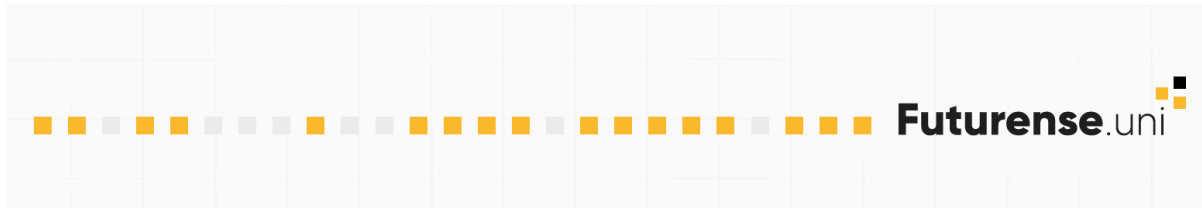
---

### Project Timeline: The Analytics Sprint

---

#### 17 **Phase 1: Mission Scoping & Hypothesis Mapping (30 minutes)**

- **Objective**: Structure your investigation and define what you're looking for.
- **Tasks**:
  1. **Launch Project Workspace in Notion AI (15 mins)**:
    - Create a new project page.
    - **Prompt for Notion AI**: "Create a project plan for a data analysis project to determine the ROI of a social media campaign. Include phases for data cleaning, integration, analysis, and reporting. Generate tasks for each phase."
    - Assign tasks to your team members within Notion.
  2. **Brainstorm Hypotheses in MindMup (15 mins)**:
    - Create a central node: "Campaign Success."
    - Branch out with your initial hypotheses. Examples: "YouTube delivers students with higher completion rates," or "Instagram has the lowest cost per signup but also the lowest engagement."
    - For each hypothesis, create sub-nodes for the data needed to prove or disprove it (e.g., "Need to join ad\_spend with completion\_percentage").



### Phase 2: The Data Laundromat (40 minutes)

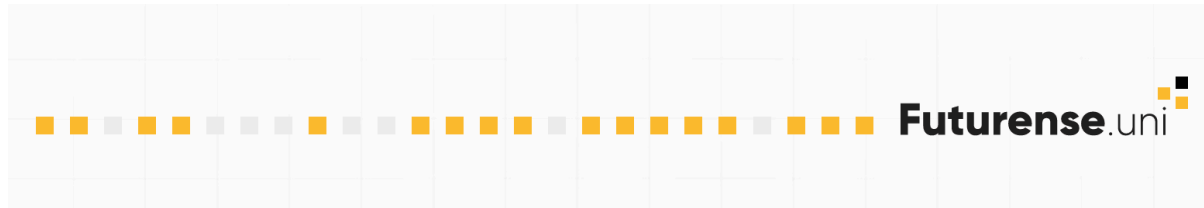
- **Objective:** Clean and standardize your raw data files.
  - **Tool: Google Colab + Python (Pandas)**
  - **Tasks:**
    1. Load `campaign_performance.csv` into a Pandas DataFrame.
    2. **Identify & Fix Messiness:**
      - **Inconsistencies:** Standardize the platform column (e.g., 'insta', 'Instagram Ads' -> 'Instagram').
      - **Duplicates:** Hunt for and remove any duplicate ad performance records.
    3. **Document Your Process:** In your Colab notebook, use Markdown cells to explain *why* you made each cleaning decision.
    4. Export the pristine data to `campaigns_cleaned.csv`.
- 

### Phase 3: The Great Merge (40 minutes)

- **Objective:** Combine marketing and student data to create a powerful master view.
  - **Tool: MySQL Workbench**
  - **Tasks:**
    1. **Set Up Your Data Warehouse:** Create a `vidyaverse_analytics` schema and import both the `student_enrollments.csv` and your newly created `campaigns_cleaned.csv`.
    2. **Write the Master Query:** Craft a SQL query using a JOIN to link the two tables. This query is the heart of your project—it connects clicks to course champions.
    3. **Validate the Merge:** Run checks to ensure the join didn't create unexpected duplicates or drop important rows.
    4. Export the final, merged data to `master_analytics_data.csv`.
- 

### Phase 4: The Eureka Moment - Analysis (40 minutes)

- **Objective:** Interrogate the master dataset to find the winning platform.
- **Tool: Google Colab + Python (Pandas)**
- **Tasks:**
  1. Load `master_analytics_data.csv` into a new DataFrame.
  2. **Engineer Key Metrics:**
    - **Cost Per Signup (CPS):** `ad_spend / total signups`.
    - **Cost Per Completed Student (CPCS):** `ad_spend / number of students with >90% completion`. This is the "killer metric."



3. **Group and Compare:** Use `groupby('platform')` to calculate the average completion\_percentage, CPS, and CPCS for each social media platform.
  4. **Synthesize Findings with Notion AI:** Paste your summary table into your Notion project page.
    - **Prompt for Notion AI:** "Summarize the key findings from this table. Which platform is the most effective and why?"
- 

### **Phase 5: Crafting the Data Story (30 minutes)**

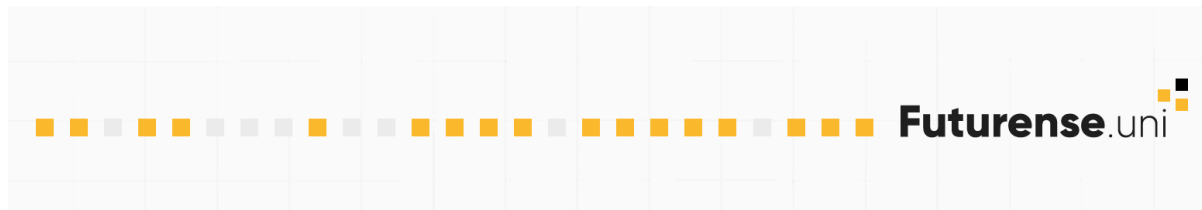
- **Objective:** Translate your numerical findings into a persuasive business recommendation.
  - **Tasks:**
    1. **Create a Presentation in Google Slides (20 mins):**
      - **Slide 1: The Punchline:** Start with your main finding (e.g., "YouTube Delivers 2x Higher ROI Despite Higher Costs").
      - **Slide 2: The Evidence:** Show the summary table comparing the platforms.
      - **Slide 3: The Recommendation:** Clearly state your advice (e.g., "Shift 30% of Instagram budget to YouTube for the next campaign").
    2. **Benchmark with Perplexity AI (10 mins):**
      - **Prompt:** "What is a good student completion rate for online self-paced courses in India?" Add this as a single line in your presentation to show you've done your market research.
- 

### **Deliverables:(In the form of project report and PPT slides)**

1. **Notion Project Page Link** (publicly shared).
  2. **MindMup Hypothesis Map** (shared link or screenshot).
  3. **Google Colab Notebook (.ipynb)** with well-commented code.
  4. **Google Slides Presentation** with your final recommendation.
- 

### **Evaluation Criteria:**

- **Analytical Rigour:** Was your data cleaning and analysis process sound?
- **Clarity of Communication:** Is your data story clear, concise, and persuasive?



- **Strategic Insight:** Did your recommendation directly address the business problem?
- **Modern Tool Usage:** Did you effectively use Notion AI and MindMap to structure your work?

#### Evaluation Criteria & Process:

- **Judged by Industry Experts:** Your analysis will be reviewed for technical accuracy and business impact.
  - **Evaluation Focus:**
    - **Technical Execution (40%):** Is your SQL and Python code clean, correct, and well-documented?
    - **Clarity of Analysis (30%):** Did you calculate the right metrics and present them clearly?
    - **Strategic Recommendation (20%):** Is your advice logical and directly supported by data?
    - **Process & Innovation (10%):** Did you use the tools effectively and show original thought?
  - 
  - **Innovation Rewarded:** Aesthetically clean presentations and innovative analytical approaches (like creating a "killer metric") will be highly valued. Plagiarized code or copied submissions will not be awarded.
  - **Bonus Points:** Awarded for consistent on-time assignment submission throughout the course.
  - **Negative points :** Awarded for every late submission from the start of submission time.
  - **Award Qualification:** Submission of at least one mini-project (FSP or BDA) is **mandatory** to be eligible for any course awards.
-