“Vidya Verse Viral Challenge – BDA Mini-Project.”

Vidya Verse Viral Challenge - Data Analysis Project Plan

Project Overview

This project aims to analyze the Return on Investment (ROI) of the VidyaVerse Viral Challenge social media campaign. Through comprehensive data analysis, we will determine the effectiveness of the campaign in terms of engagement, conversion, and financial returns.

Project Phases

1. Data Collection Phase

- Data sets are already provided and were used.

2. Data Cleaning Phase

- **Remove duplicate entries** across all dataset
- **Handle missing values** with appropriate statistical methods
- **Standardize data formats** across different platforms
- **Normalize time zones** for consistent temporal analysis
- **Filter out bot activity** and non-genuine engagements
- **Validate data integrity** through cross-referencing sources

3. Data Integration Phase

- **Create unified data schema** for cross-platform analysis
- **Develop user journey mapping** from social media interaction to conversion
- **Link engagement metrics** with conversion data
- **Integrate cost data** with performance metrics
- **Create master dataset** with appropriate relational structure

4. Data Analysis Phase

- **Calculate key performance indicators** (KPIs) for the campaign
- **Perform trend analysis** of engagement over time
- **Conduct segmentation analysis** based on demographics and behavior
- **Calculate direct ROI** using revenue attribution models
- **Perform statistical testing** to validate findings

- **Apply predictive modeling** for future campaign optimization
- **Conduct comparative analysis** with industry benchmarks

5. Reporting Phase

- **Create executive summary** with key findings and ROI metrics
- **Develop interactive dashboard** for stakeholder exploration
- **Generate detailed technical report** with methodology and findings
- **Prepare data visualizations** highlighting key insights
- **Draft recommendations** for future campaign optimization
- **Present findings** to key stakeholders

Timeline and Milestones

Phase	Duration	Key Deliverables
Data Collection	1 days	Raw data repository were already provided
Data Cleaning	2 days	Cleaned datasets, Data quality report
Data Integration	2 days	Unified database, Integration documentation
Data Analysis	2 days	Analysis notebooks, Statistical reports
Reporting	1 days	Executive dashboard, Technical report, Presentation

Resources Required

- **Personnel:** Data Analyst, Database Administrator, Business Analyst, Data Visualization Specialist
- **Software:** SQL Database, Python for analysis, notion AI , Google slides , google docs , google sheets , perplexity , chat-gpt

ROI Calculation Methodology

ROI will be calculated using the following formula:

```
ROI = ((Revenue attributed to campaign - Campaign costs) / Campaign costs) × 100%
```

Risk Management

Risk	Mitigation Strategy
Data privacy concerns	Ensure compliance with data protection regulations
Attribution accuracy	Implement multiple attribution models for comparison
Timeline delays	Build buffer time into each phase

Success Criteria

- Identification of top-performing content and channels
- Actionable recommendations for future campaigns
- Interactive dashboard for ongoing performance monitoring

Project Task Table

Title	Due Date	Owner	Phase	Status
Set up Notion workspace & run AI project plan prompt	26 august	krishna	Phase 1 – Scoping/mapping	completed
Assign tasks to team in Notion	26 august	krishna	Phase 1 – Scoping	completed
Map hypotheses & required data in MindMup	26 august	ishita	Phase 1 – Scoping	completed
Clean raw CSVs & standardize values	26 august	samiksha	Phase 2 – Data Cleaning	completed
Document cleaning steps in Notion	26 august	krishna	Phase 2 – Data Cleaning	completed
Export cleaned CSV	27 august	samiksha	Phase 2 – Data Cleaning	completed
Import cleaned CSVs into MySQL schema	27 august	samiksha	Phase 3 – Data Integration	completed
Merge campaign & results tables with JOIN	27 august	samiksha	Phase 3 – Data Integration	completed
Validate merged data (no duplicates/missing)	27 august	samiksha	Phase 3 – Data Integration	completed
Calculate CPS & CPCS in Colab	27 august	samiksha	Phase 4 – Analysis	completed
Build summary table by platform	27 august	samiksha	Phase 4 – Analysis	completed
Summarize findings with Notion AI	27 august	krishna	Phase 4 – Analysis	completed
Draft slides (Punchline, Evidence, Recommendation)	28 august	suresh	Phase 5 – Reporting	completed
Publish deliverables (Notion page, MindMup, CSV, SQL, Colab, Slides)	29 august	team	Deliverables	completed
Add benchmark line from Perplexity research	28 august	ishita	Phase 5 – Reporting	completed

After completion of all the tasks the final observations are as follows…

Synthesize Findings with Notion AI

Name	average_completion_percentage	average_cpcs	average_cps	std_dev_completion_percentage	std_dev_cps	total_ad_spend
Instagram	58.14	inf	4577.322005772006	26.198574327131	4913.121000868606	192817.08000000002
LinkedIn	56.303513513514	inf	4172.811265811966	25.436750214183	2762.738904961672	273447.22000000003
YouTube	50.561428571429	inf	4515.266446825397	27.237625178154	3811.572294228121	242526.21

Database Summary: Platform Performance Analysis

Database Structure

This database contains **3 social media platforms** with comprehensive performance metrics including:

- **Completion rates** and their variability
- **Cost-per-share (CPS)** metrics and standard deviations
- **Total advertising spend** per platform
- **Cost-per-completed-share (CPCS)** data

Key Metrics Overview

Completion Performance:

- Range: 50.56% to 58.14% completion rates
- Instagram leads with highest completion (58.14%)
- Moderate variability across platforms (25-27% standard deviation)

Cost Structure:

- Total spend range: ₹192K to ₹273K
- Combined total investment: ₹708,790 across all platforms
- Cost-per-share varies from ₹4,173 to ₹4,577

Performance Consistency:

- All platforms show similar volatility patterns
- LinkedIn demonstrates most consistent cost-per-share (lowest std dev: ₹2,763)
- Instagram shows highest CPS variability (std dev: ₹4,913)

Data Quality Notes

- Complete dataset with no missing values

- All cost-per-completed-share values marked as "inf" suggesting either infinite costs or measurement limitations
- Metrics appear to be campaign averages rather than time-series data

This database serves as the foundation for ROI calculations and platform effectiveness analysis in your viral challenge campaign assessment.