DATA ANALYTICS FEATURES FOR AN E-LEARNING PLATFORM

INTERNSHIP ASSIGNMENT – UTKARSH CHOUHAN

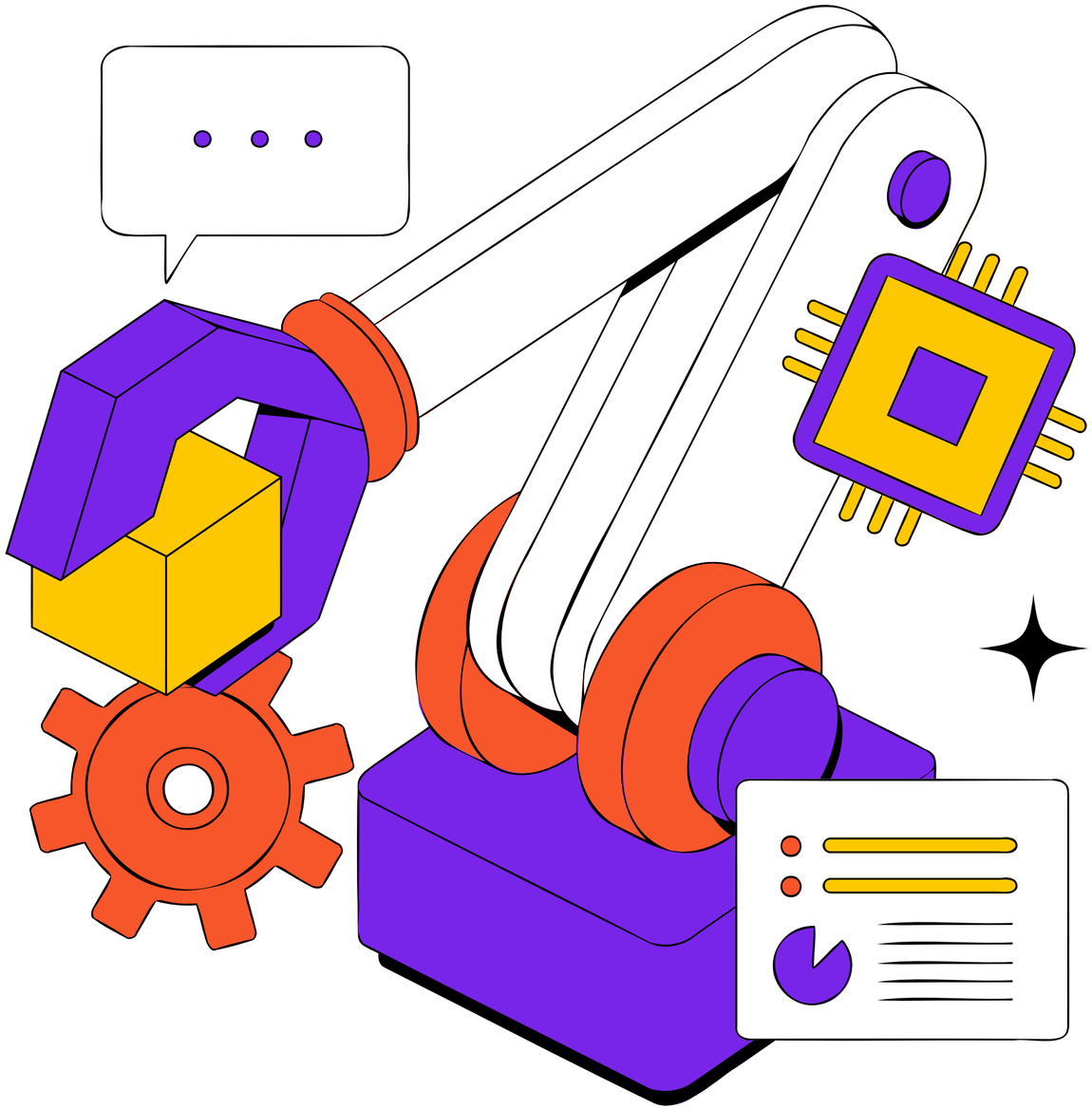
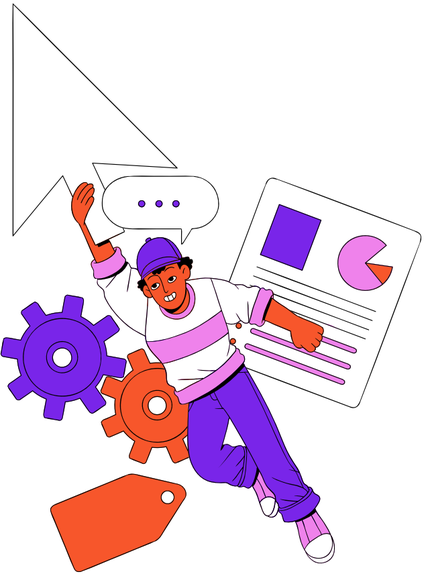


TABLE OF CONTENTS

01. ANALYTICAL FEATURES

- **Student Drop-off/Churn Predictor:** Analyzing at which second/minute of a video students stop watching to identify boring content.
- **Course Completion Probability Score:** A dynamic score assigned to every user based on their interaction frequency, predicting if they will finish a course.
- **Instructor "Engagement" Heatmap:** Visualizing which topics in a syllabus get the most rewinds (high interest/confusion) vs. skips (low value).
- **Upsell Recommendation Engine:** Analyzing "Quiz Scores" to recommend the next course (e.g., High score = Advanced course; Low score = Remedial course).
- **Search vs. Content Gap Analysis:** Tracking what users search for vs. what courses actually exist, to identify market demand for new topics.



02. THE DEEP DIVE & LOGIC

Feature 1: The "At-Risk" Student Predictor (Churn Predictor).

- This is the strongest choice for analysis because it directly impacts revenue.
- Companies care most about keeping customers they already have.

LOGIC:

We are building a rule-based engine that flags students as "High Risk" if they exhibit specific withdrawal behaviors. The logic checks two main indicators:

1. Inactivity: Have they stopped logging in?
2. Stagnation: Are they logging in but failing to make progress (low watch time or failing quizzes)?

PSEUDOCODE

```
INPUT: Student_ID, Days_Inactive, Video_Completion_%, Avg_Quiz_Score

# Define Business Thresholds
SET MAX_INACTIVITY_DAYS = 7
SET MIN_COMPLETION_RATE = 20%
SET MIN_PASSING_SCORE = 50%

FOR EACH Student IN Database:

    # Rule 1: Check for "Ghost" Users (High Inactivity)
    IF Days_Inactive > MAX_INACTIVITY_DAYS THEN
        SET Status = "High Risk (Absent)"
        TRIGGER Action = "Send 'We Miss You' Email"

    # Rule 2: Check for "Stuck" Users (Low Performance)
    ELSE IF Video_Completion_% < MIN_COMPLETION_RATE OR Avg_Quiz_Score < MIN_PASSING_SCORE THEN
        SET Status = "Medium Risk (Struggling)"
        TRIGGER Action = "Recommend Easy Prerequisite Course"

    # Rule 3: Healthy Users
    ELSE
        SET Status = "Low Risk (Healthy)"
        TRIGGER Action = "None"

    END IF

    PRINT Student_ID, Status, Action

END LOOP
```

EXPLAINATION OF PSEUDOCODE

- **Input Processing**: The logic begins by iterating through the database to analyze key behavioral metrics for every single student: their last login date (Days_Inactive), content consumption (Video_Completion_%), and academic performance (Avg_Quiz_Score).

- **Defining Success Metrics**: It establishes strict "Business Thresholds" to serve as benchmarks: a maximum of 7 days for inactivity, a minimum of 20% for video completion, and a passing grade of 50% for quizzes.

- **Priority Check - "Ghost" Users (Rule 1)**: The system first checks for retention risk. Any student who hasn't logged in for more than 7 days is immediately flagged as "High Risk (Absent)", triggering an automated "We Miss You" email to bring them back.

EXPLAINATION OF PSEUDOCODE

- **Performance Check - "Struggling" Users (Rule 2)**: If a student is active but falling behind (low video completion) or failing tests (low quiz scores), they are classified as "Medium Risk (Struggling)". The system helps them by recommending an easier prerequisite course.

- **Validation - "Healthy" Users (Rule 3)**: Students who pass both the activity check (logging in weekly) and the performance check (good grades/progress) are marked as "Low Risk", indicating they are on a successful learning path requiring no intervention.

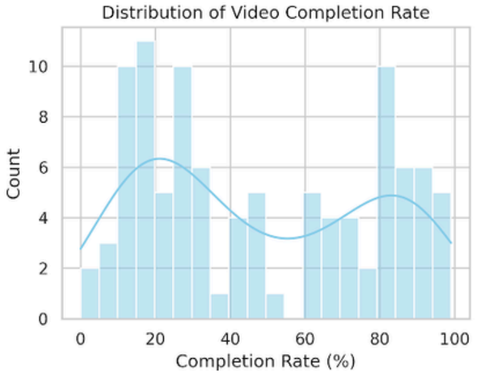
- **Actionable Output**: The loop concludes by outputting a specific status and a distinct action item for every Student_ID, effectively automating the platform's retention and support strategy.

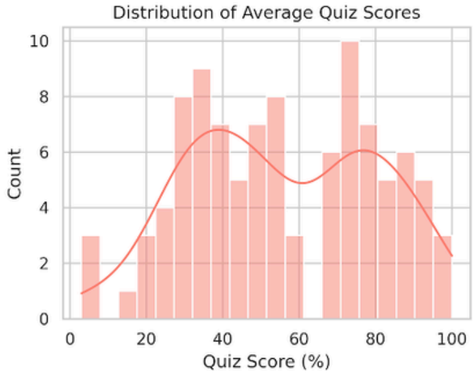
DATASET

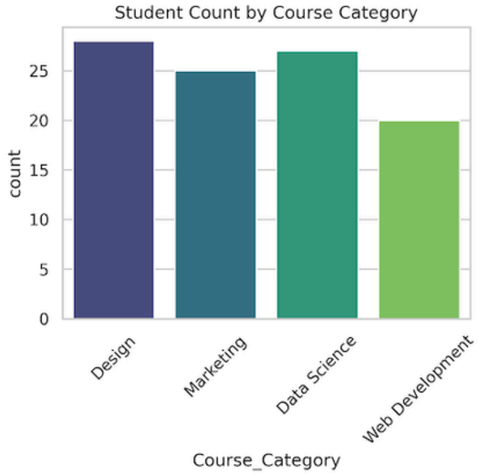
Data Dictionary: At-Risk Student Predictor Dataset

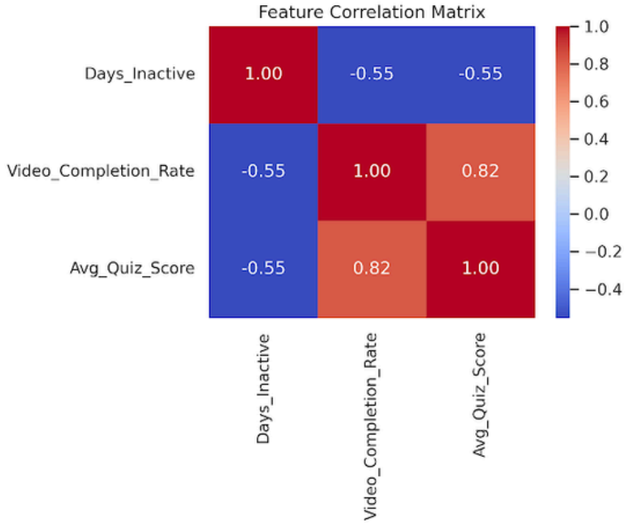
Feature Name	Description	Business Relevance
Student_ID	Unique identifier for each student.	Tracking individual user journeys.
Days_Inactive	Number of days since the student last logged in.	Primary indicator of churn risk (Recency).
Video_Completion_Rate	Percentage (0-100) of video content watched.	Measures engagement and content stickiness.
Avg_Quiz_Score	Average score (0-100) across all quizzes.	Indicates comprehension and potential frustration.
Course_Category	Subject of the course (e.g., Data Science, Design).	Identifies trends in subject popularity/difficulty.
Subscription_Status	Current status: 'Active' or 'Cancelled'.	Target variable for revenue analysis.
Risk_Segment	Derived label: 'Healthy', 'Struggling', or 'High Risk'.	Actionable segments for marketing intervention.

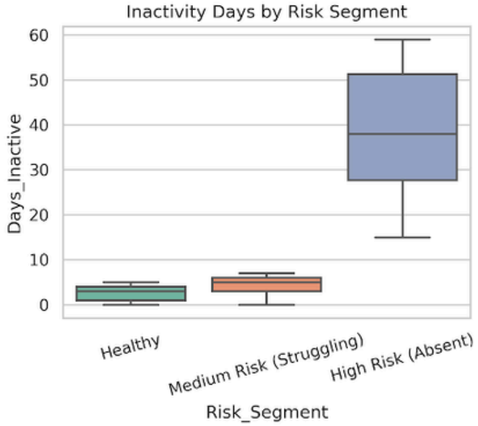
03. DATASET: ANALYSIS AND PLOTS

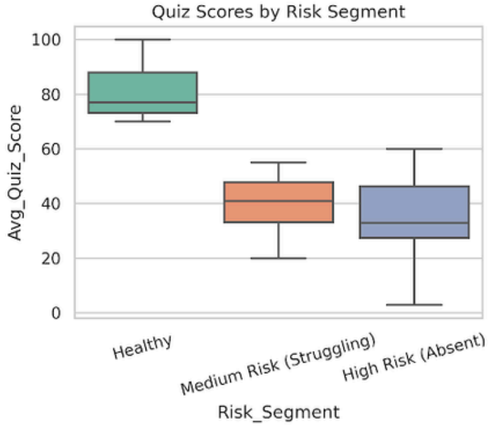








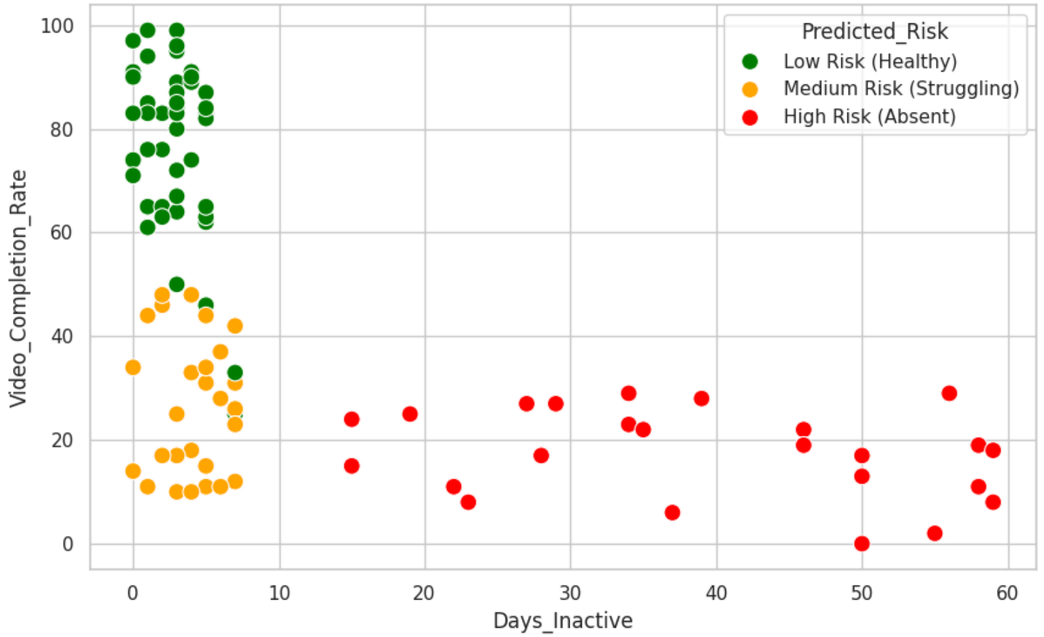




03. DATASET: ANALYSIS AND PLOTS

1. Final Results (The "What")

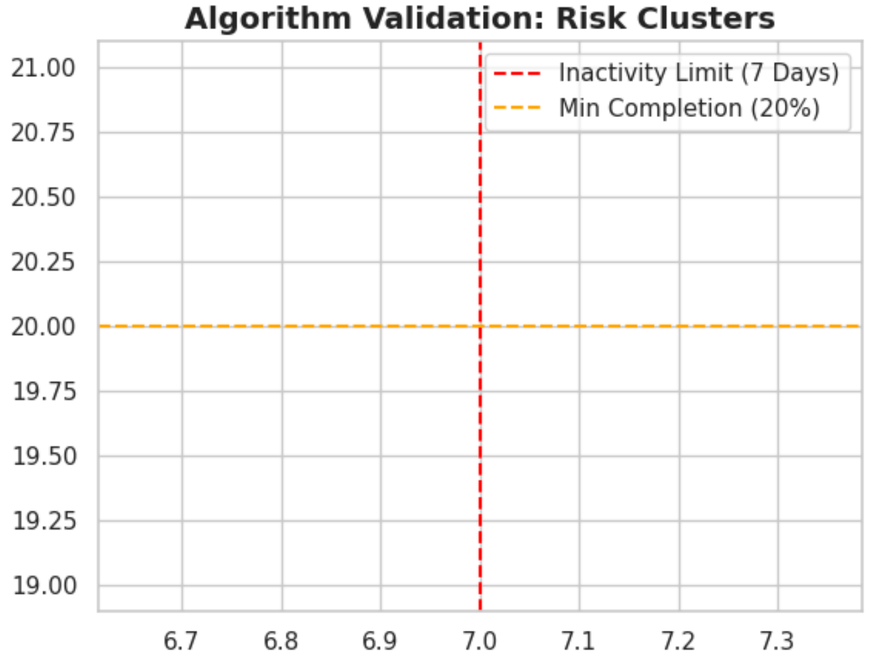
- Healthy Users (46%): Nearly half of your students are on track and require no intervention.
- Struggling Users (30%): A significant portion are logging in but failing to progress due to difficulty.
- High-Risk / Absent Users (24%): Nearly one-quarter of students have effectively abandoned the platform (inactive > 7 days).



03. DATASET: ANALYSIS AND PLOTS

2. Key Observations (The "Why")

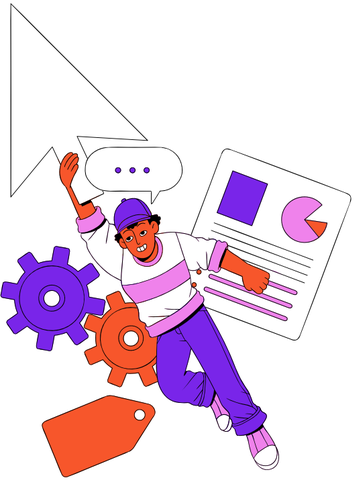
- The "37-Day" Gap: Students who cancelled their subscriptions had an average inactivity of 37.2 days, compared to just 7.9 days for active users. This proves that long gaps in login history are the #1 predictor of churn.
- The Difficulty Wall: The "Struggling" segment has an average quiz score of 38.5%, compared to 78.3% for healthy users. This suggests these students aren't lazy; they are stuck.
- Category Risk: Data Science and Web Development courses have the highest number of "Absent" students (7 each), likely due to the steeper learning curve compared to Design.



CONCLUSION AND IMPLEMENTATION ROADMAP

- STUDENT CHURN ON OUR PLATFORM IS NOT RANDOM; IT IS HIGHLY PREDICTABLE. WE LOSE 24% OF REVENUE NOT BECAUSE STUDENTS 'LOSE INTEREST', BUT BECAUSE WE FAIL TO INTERVENE DURING THE CRITICAL 7-14 DAY WINDOW OF INACTIVITY OR WHEN THEIR QUIZ SCORES DROP BELOW 40%.

Feature to Implement	Why (Insight)	Action (Solution)
1. The 'Day 7' Auto-Nudge	Analysis shows inactivity > 7 days leads to churn.	Send 'Come back' email exactly 7 days after last login. Prevents the '37-day gap'.
2. Remedial Content Suggestions	30% of users are 'Struggling' (Avg Score: 38.5%).	If quiz < 50%, block progress. Pop up 'Review this 5-min video' suggestion.
3. Difficulty Audits	Data Science has the highest drop-off rate.	Review Week 2 & 3 Data Science videos. Material likely too hard or poorly explained.
4. Instructor Dashboards	Instructors currently fly blind.	Show 'At Risk' students to instructors. Personal notes increase retention by ~20%.



TURNING INSIGHTS INTO REVENUE: A 3-POINT RETENTION STRATEGY

1. AUTOMATE RE-ENGAGEMENT (TARGETING THE 24% GHOSTS).
2. SUPPORT THE STRUGGLERS (TARGETING THE 30% STUCK USERS).
3. OPTIMIZE CONTENT (TARGETING THE HIGH-RISK CATEGORIES).

CODES AND DATA

DRIVE LINK WITH PROJECT FOLDER AND DATA:

HTTPS://DRIVE.GOOGLE.COM/DRIVE/FOLDERS/1ZNLLFTOAAJKX_YIXTCNXOF0NA800_XKB?USP=DRIVE_LINK

THANK YOU!