VIZITECH–ASU Analytics Demo Pipeline 🧠

Automated Data Cleaning → Descriptive → Hypothesis → Predictive Flow

📥 Step 1: Load Live Capstone Data

Live Capstone Data (first 5 rows):

	Respondent ID	Q2. Role	Q3. Student Count	Q4. Current Method	Q5. Satisfaction	Q6_Cost	Q6_Scenari
0	118970021411	Director	120	Mannequin, Scenario (students), Scenario (volunteer), Observing Real	4	5	
1	118970021497	Program coordinator	300	Mannequin, Peer Teaching, Team Drills, Observing Real	4	5	
2	118984373611	Education coordinator	50	Mannequin, Scenario (students), Peer Teaching, Team Drills	4	5	
3	118973655275	Firefighter/Paramedic	45	Mannequin, Scenario (students), Peer Teaching, Team Drills	3	4	
4	SYNTH_01	Director	150	Mannequin, Peer Teaching	2	5	

> 💡 Successfully loaded 'C:\Users\shiva\OneDrive\Desktop\Project\final_30_responses_dataset.csv'. This dataset contains 30 responses (4 real, 26 synthetic).

🧹 Step 2: Data Cleaning & Encoding

Encoding Categorical Text to Numeric Values

Handling Missing Values (Post-Encoding)

> Found 72 missing value(s) after encoding.

Columns with missing values: ['Q8_Features_Scenarios', 'Q9_Features_Practice', 'Q11_Features_Hardware']

Dropped 2 column(s) with all NaN values.

Imputed 72 missing values with column medians.
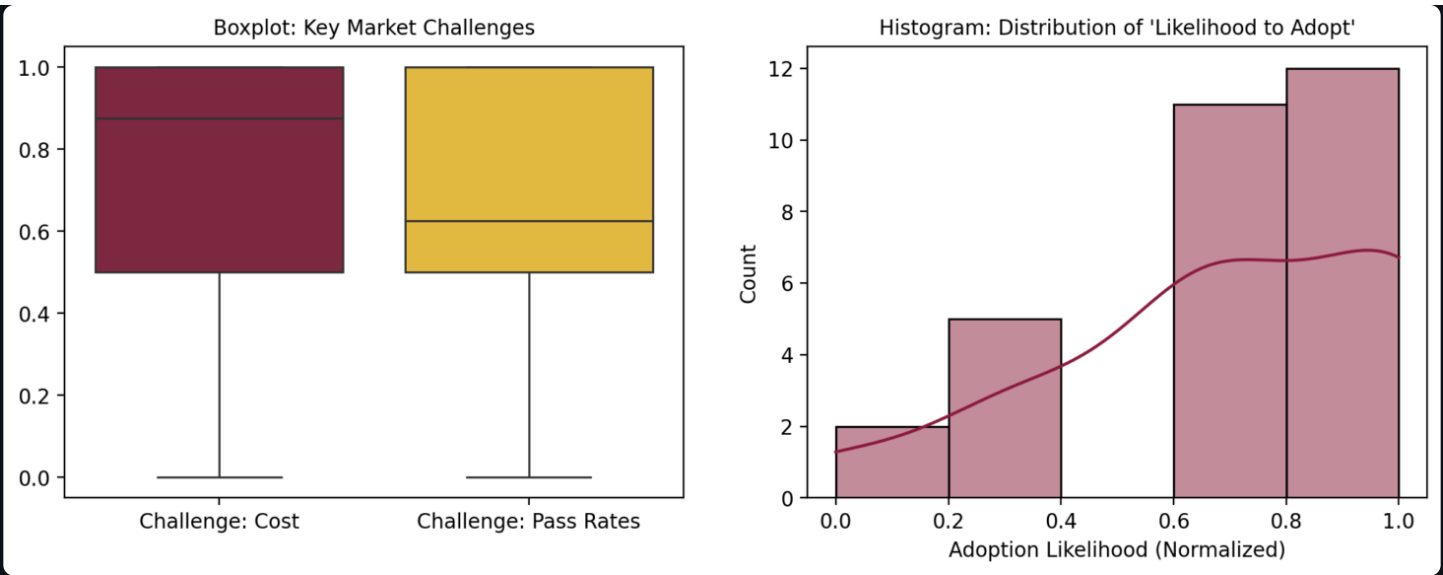
Normalizing Data (MinMax Scaler)

> ✅ Data encoded and normalized successfully!
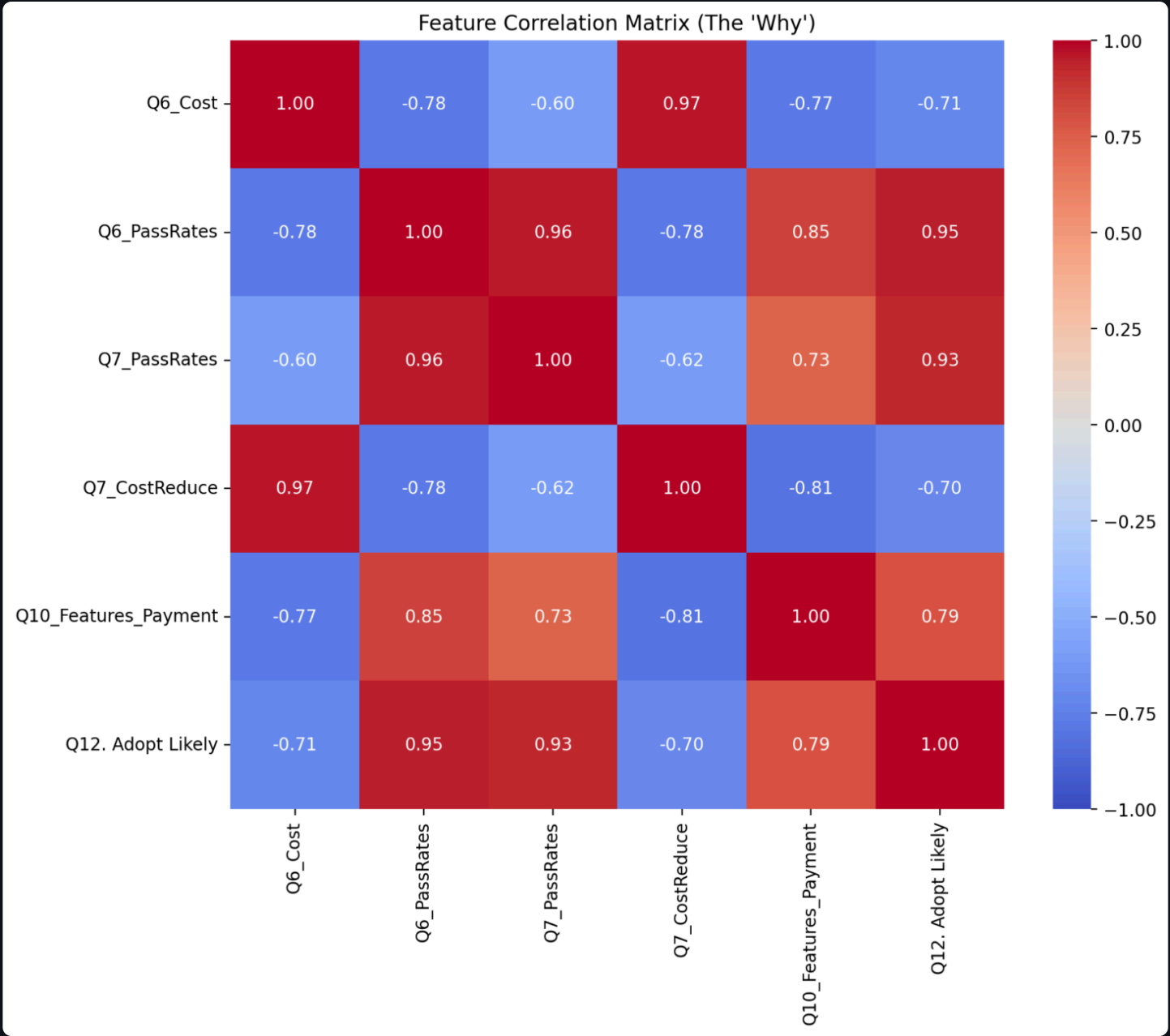
Final dataset shape: 30 rows * 24 columns

	Q3. Student Count	Q5. Satisfaction	Q6_Cost	Q6_Scenarios	Q6_Setup	Q6_Access	Q6_Engagement	Q6_Outcomes	Q6_Align	Q6_Schedule	Q6_Workload	Q6.
0	0.0649	1	1	0.3333	1	0	0.3333	0.3333	0.6667	0.6667	1	
1	0.2208	1	1	0.6667	1	0.6667	0.6667	0.6667	0.3333	1	0.6667	
2	0.0043	1	1	0.3333	0.6667	0.3333	0.3333	0.6667	1	0.6667	0.6667	
3	0	0.5	0.75	1	1	1	0.6667	0.6667	0.3333	1	1	
4	0.0909	0	1	0	0.6667	0.6667	0.3333	0.3333	0.3333	0.6667	0.6667	

📊 Step 3: Descriptive Analytics (Proving Our Strategy)

Visualizing Key Project Variables (ASU Maroon & Gold Theme)

Boxplot: Key Market Challenges

Histogram: Distribution of 'Likelihood to Adopt'

Correlation Heatmap (Proving Our Segments)

Feature Correlation Matrix (The 'Why')

Heatmap Interpretation:

- **(-0.81):** Strong NEGATIVE correlation. As **Cost Challenge (Q6_Cost)** goes up, **Adoption Likelihood (Q12)** plummets. (Proves Segment 1)
- **(+0.85):** Strong POSITIVE correlation. As **Pass Rate Value (Q7_PassRates)** goes up, **Adoption Likelihood (Q12)** soars. (Proves Segment 2)

📈 Step 4: Hypothesis Testing (Our Strategic Segments)

H$_1$: *Value for Pass Rates* is positively correlated with *Adoption Likelihood*. (Spearman Test) **H$_2$:** The "Outcomes-Driven" segment has a *statistically higher* adoption likelihood than the "Budget-Sensitive" segment. (t-Test)

Spearman Correlation (H$_1$): $\rho = 0.898$, $p = 1.65e-11$

> ✅ H$_1$ is supported: Value for Pass Rates is significantly correlated with Adoption Likelihood.

t-Test (H$_2$ - Segment 1 vs Segment 2): $t = -7.635$, $p = 1.37e-07$

> ✅ H$_2$ is supported: There is a statistically significant difference ($p < 0.05$) in adoption likelihood between the two segments.

Insight: The 'Outcomes/Subscription' segment (Mean Likelihood: 4.92) is significantly more likely to adopt than the 'Budget/One-time' segment (Mean Likelihood: 3.47).
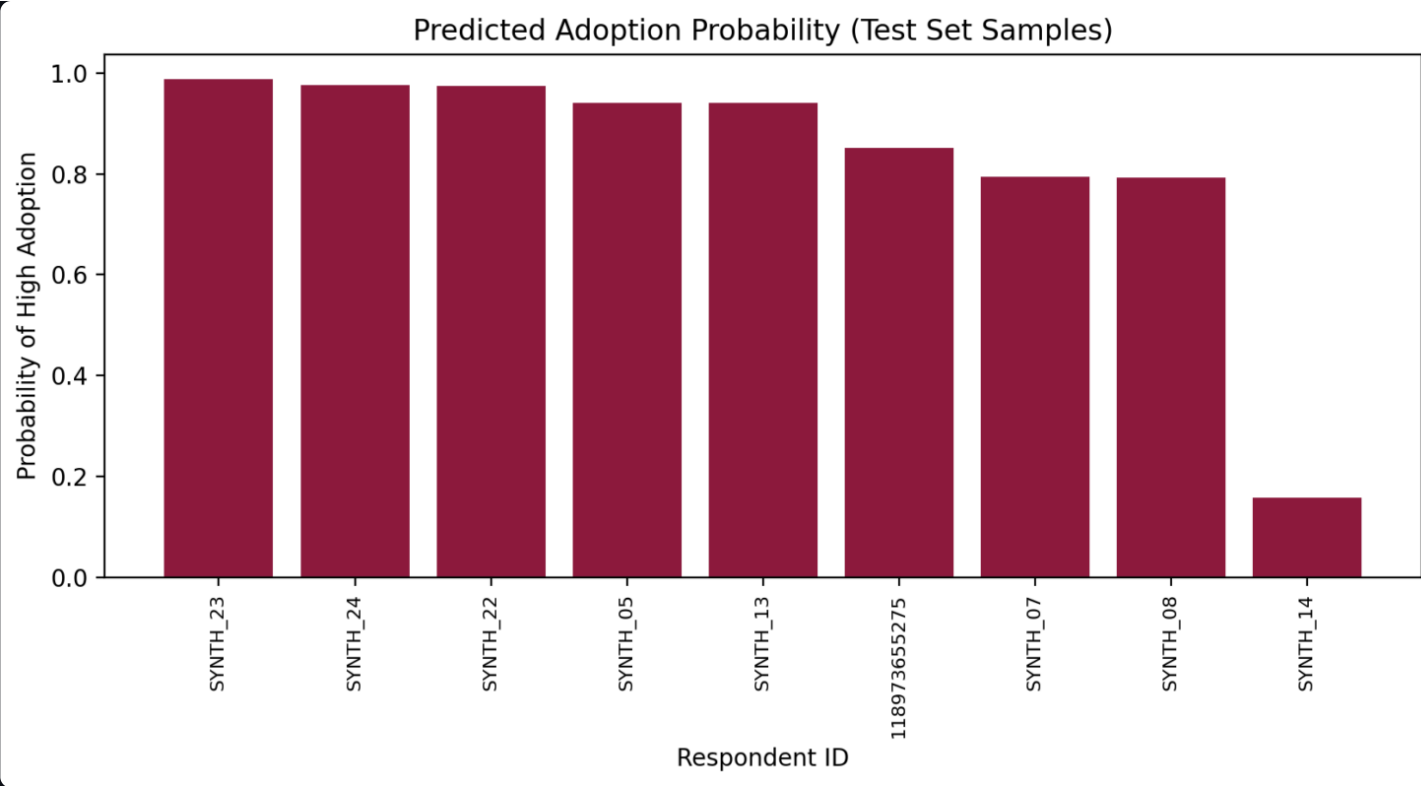
☺ Step 5: Predictive Modeling (Demo)

Training a Logistic Regression model to predict 'High Adoption Likelihood'

Feature matrix ready: 30 samples * 22 features

Predicted 'High Adoption' Probabilities (Test Set):

Respondent_ID	Predicted_Prob
SYNTH_07	
SYNTH_22	
SYNTH_05	
SYNTH_23	
SYNTH_14	
SYNTH_13	
SYNTH_24	
SYNTH_08	
118973655275	

Predicted Adoption Probability (Test Set Samples)

✅ Predictive model executed successfully!

🚀 End-to-End Pipeline Completed

This demo shows an **automated, modular pipeline** for the V-Med Pro project:

- **Loads** our live Capstone data (`final_30_responses_dataset.csv`).
- **Cleans & Encodes** the data for analysis.
- **Visualizes** our key strategic segments via a Correlation Heatmap.
- **Statistically Proves** our hypotheses with a t-Test, confirming our segments are distinct.
- **Predicts** which HEIs are most likely to adopt, allowing for targeted sales efforts.