

Analytic Process To Produce Visualization

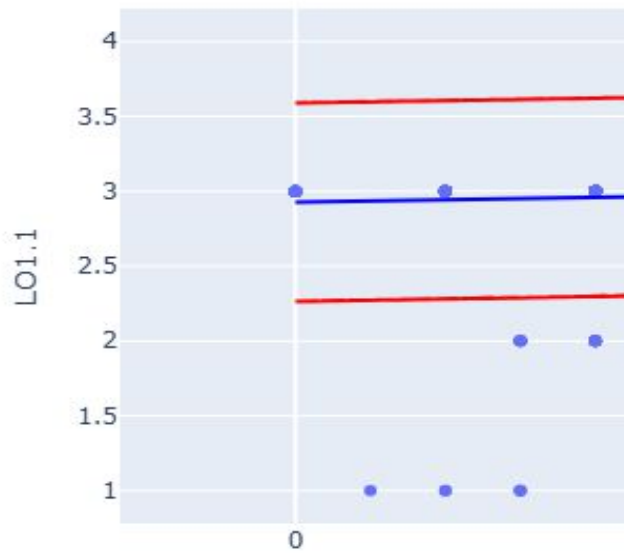
Use this workbook to describe one single analytic process or a

Category of Design Component	Data Fields
Identification	Process / Analysis Name
Identification	Requirement(s) Satisfied
Identification	Process Description
Source	Data Source
Destination	Data Destination Database
Destination	ER Diagram Subset Loaded
Transformation	Data Transformation Required
Transformation	Pre-processing Required
Display	Type of Graphic Display
Display	Example Screenshot
Display	User-Facing Description
Display	Filters / Parameters
Display	Measures / Dimensions
Operations	Refresh Frequency

Operations	Dependencies / Assumptions
Operations	Additional Notes

Screenshot Area

Regression: LO1.1 ~ LO1



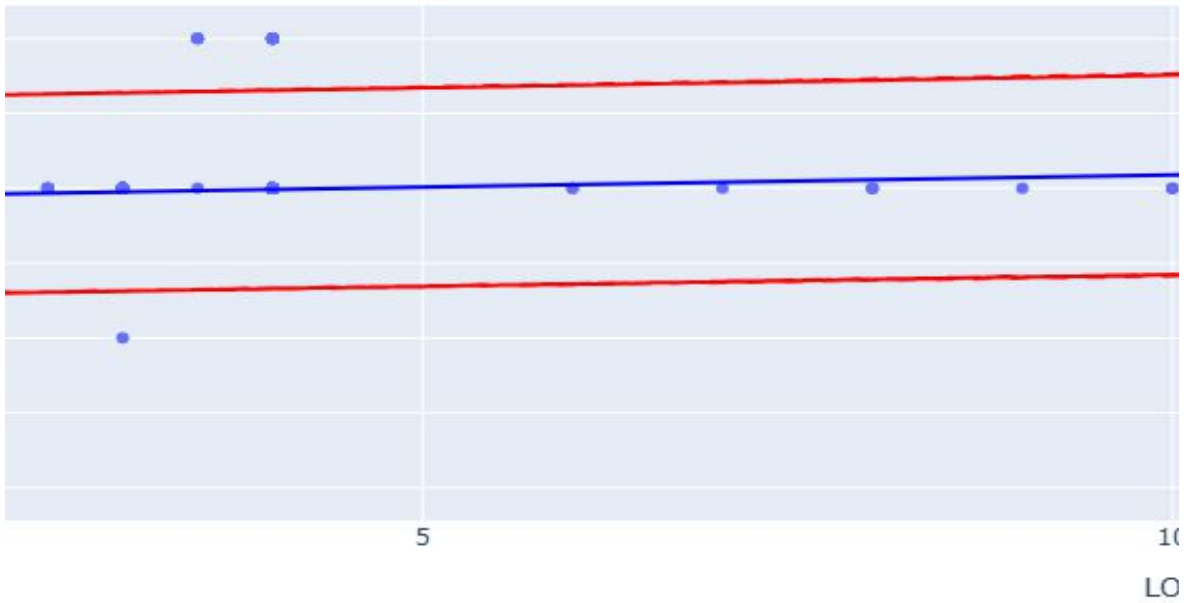
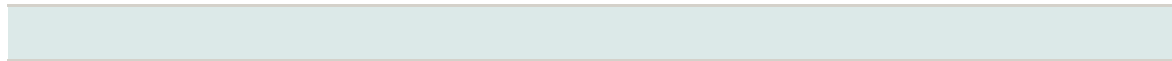
tion

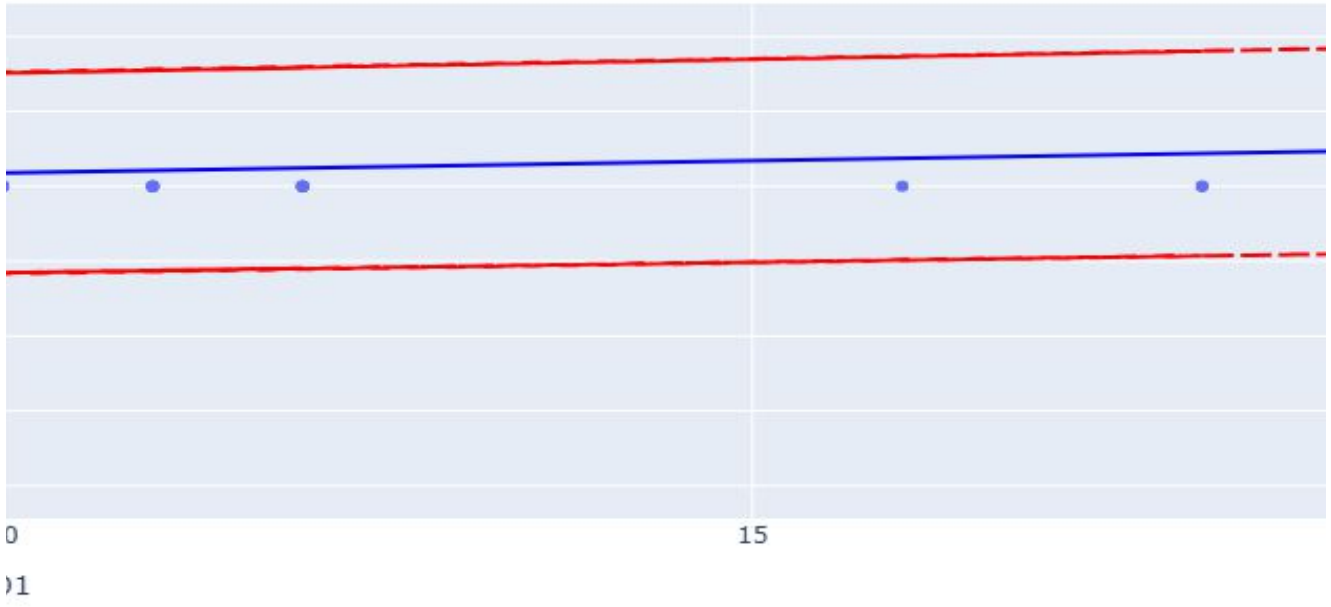
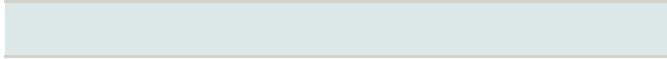
analysis that results in a graph or chart displayed to the user.

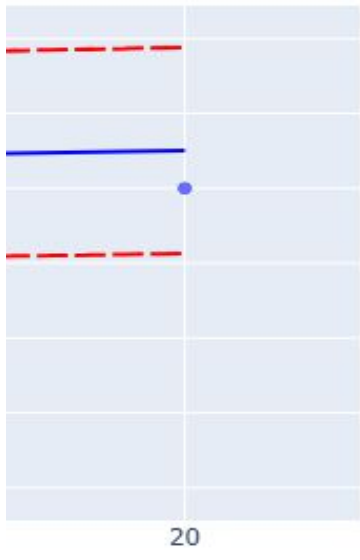
Actions Programmed/Script to Run/Tools Used
GE Data Regression & Cluster Analysis
Visualize relationships between numeric GE assessment variables; identify student groupings via unsupervised clustering
Loads raw GE assessment records from Supabase, applies KMeans clustering on the first 4 numeric columns, and renders either an OLS regression scatter (with 95% prediction intervals) or a cluster-colored scatter plot
Supabase (PostgreSQL) — table GEData_Raw_2021_2024 via REST API
In-memory pandas DataFrame (no persistent warehouse; live query on app start)
Single flat table: GEData_Raw_2021_2024; no joins — fields include Subject, Goal, Delivery, LO1–LO5, and other numeric assessment columns
Coerce all columns to numeric where possible; drop rows where all numeric cols are null; fill remaining nulls with column median; append KMeans cluster label (3 clusters)
Column-type detection (numeric vs. categorical); median imputation for missing values; cluster assignment via KMeans (n=3, n_init=10) on first 4 numeric columns
Scatter plot — either OLS regression line with 95% prediction interval bands, or cluster-colored scatter
N/A
Users select X and Y axes from any numeric column and toggle between Regression view (showing trend line + confidence bands) and Cluster view (color-coded by auto-detected grouping). Filterable by cluster and Subject.
X-axis column, Y-axis column, Graph type (Regression / Cluster), Cluster filter (multi-select), Subject filter
X = selected numeric column, Y = selected numeric column, Color = cluster (cluster view), Hover = categorical string columns
On-demand (data loaded once at app startup from Supabase)

Supabase table must be accessible; at least 2 numeric columns required for clustering; OLS regression falls back to plain scatter if model fitting fails

Reset button restores all filters to defaults. KMeans clustering uses the first 4 numeric columns detected at runtime.







— OLS Fit
- - 95% PI Lower
- - 95% PI Upper