



# Kamya Sarda

Data Viz Portfolio

2025

# Introduction

A data mining for business intelligence project that involves picking real-world datasets and making visualisations using **Tableau** that tell a story in business contexts, and build & compare models using **RStudio**.



# Student Depression Dataset (Logistic Regression)

**Objective: Predict likelihood of depression based on numeric factors.**

Primary Research Question: "Which numeric factors such as academic pressure, CGPA, and study satisfaction significantly predict depression in students?"

Source: Kaggle

## Key Variables:

- Academic Pressure
- Study Satisfaction
- CGPA
- Other factors like Job Satisfaction, Age, etc

## EDA Highlights (Tableau)

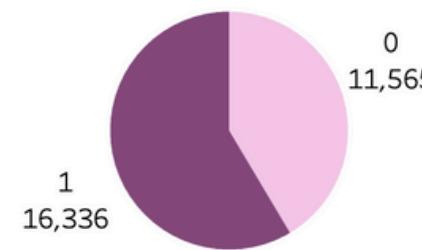
- Higher Academic Pressure linked to reduced sleep.
- Financial stress and suicidal ideation strongly associated with depression.
- Males reported slightly higher depression levels.

# Dashboard 1

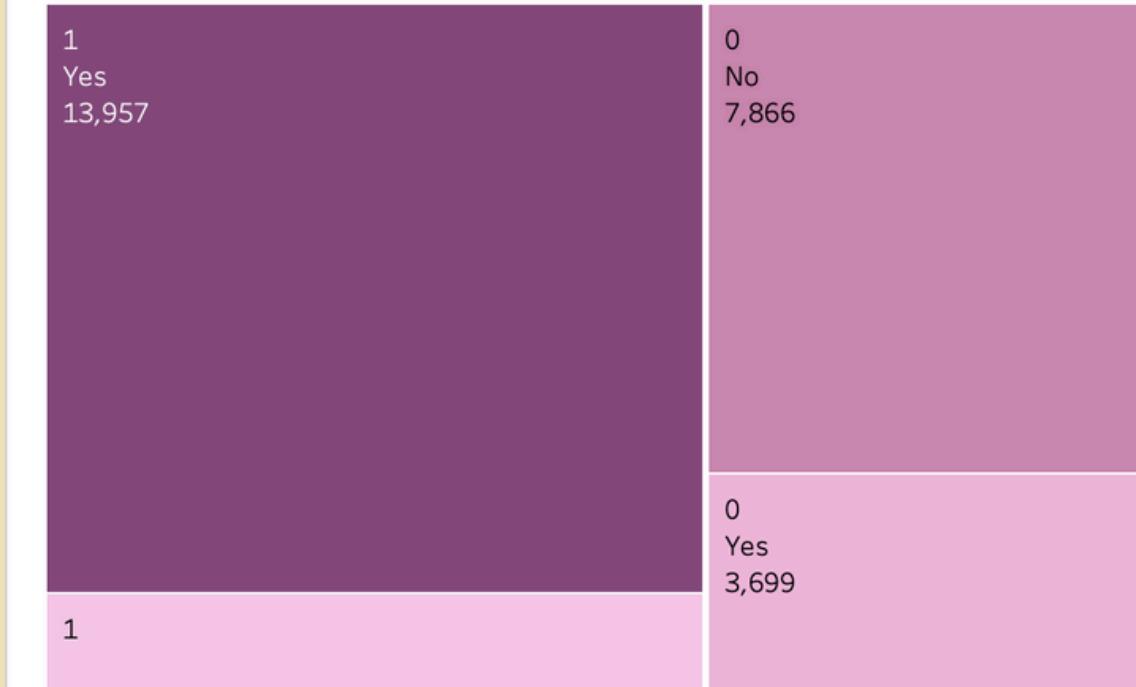
STUDENT MENTAL  
HEALTH INSIGHTS  
BASED ON SLEEP  
DURATION, SUICIDAL  
THOUGHTS, FINANCIAL  
STRESS:

## Student Mental Health Insights:

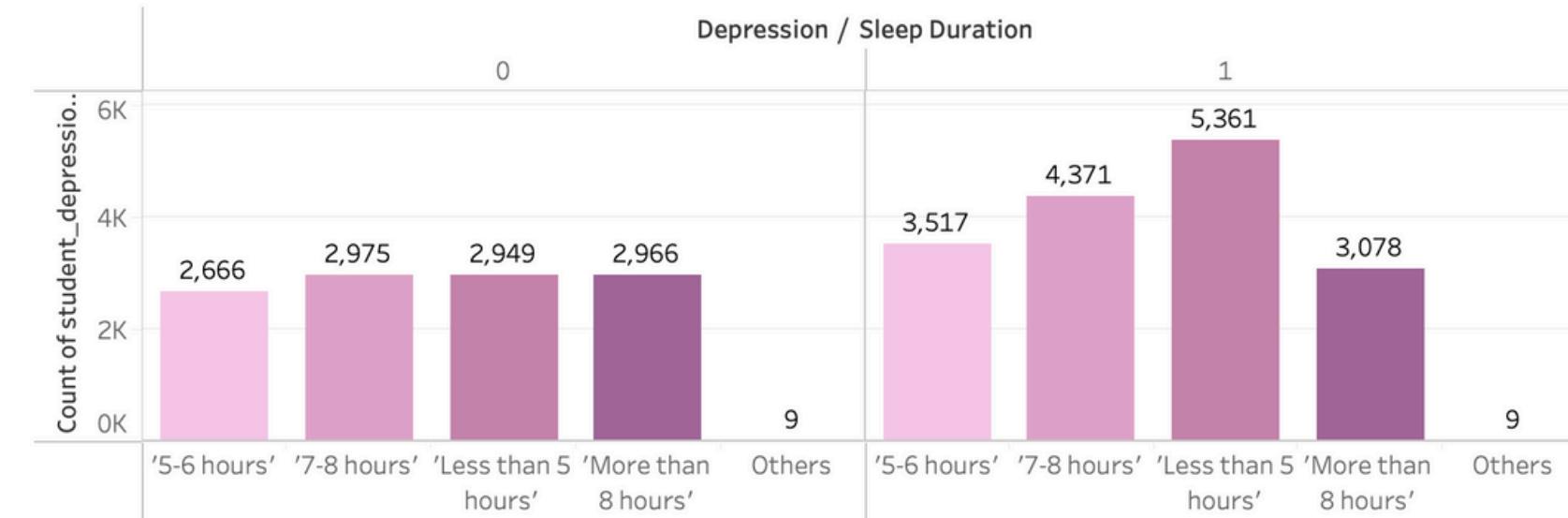
### Proportion of Depression



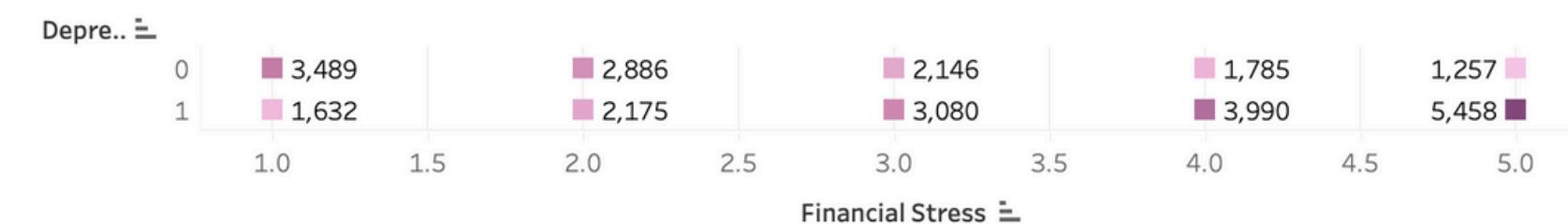
### Suicidal Thoughts vs Depression



### Sleep Duration vs Depression



### Financial Stress vs Depression



### Sleep Duration

Number of Students

- '5-6 hours'
- 'Less than 5 ho..'
- '7-8 hours'
- 'More than 8 ho..'
- Others

### Number of Students

1,257 5,458

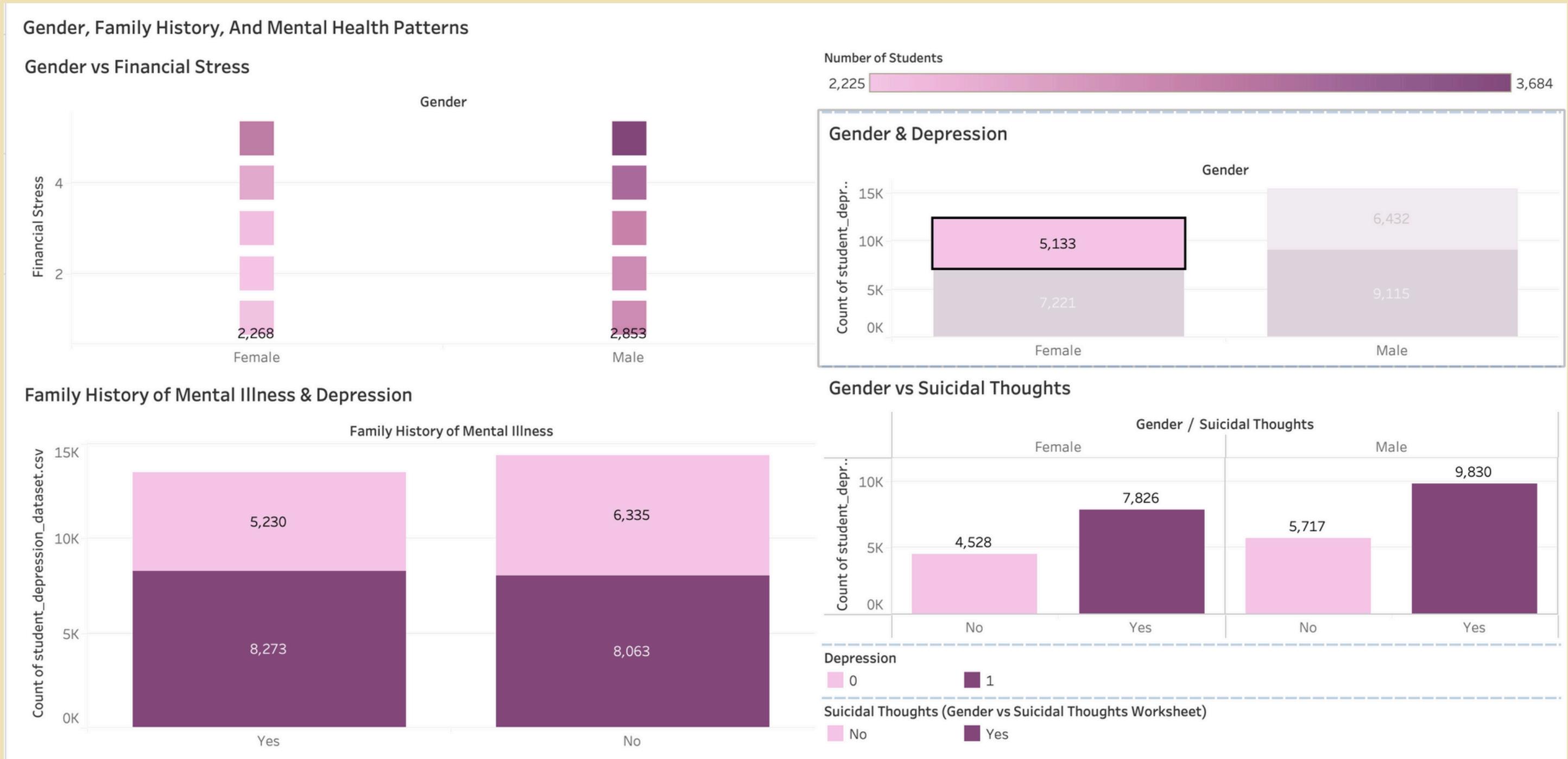
# Dashboard 2

## STUDENT MENTAL HEALTH INSIGHTS BASED ON ACADEMIC STRESS, AGE, STUDY SATISFACTION AND DEPRESSION SCORE



# Dashboard 3

## STUDENT MENTAL HEALTH INSIGHTS BASED ON GENDER & FAMILY HISTORY



# Model Building: RStudio

## Method: Logistic Regression

- Final Model (post Stepwise AIC selection):
  - Academic Pressure (+)
  - Study Satisfaction (-)
  - CGPA (-)
- Performance:
  - Pseudo R-squared: ~30%
  - Accuracy: ~75–80%

## Graphs from RStudio Analysis (Diagnostics):

- ROC Curve: High AUC (Area Under Curve) – Indicates good model performance.
- Confusion Matrix: Balanced sensitivity and specificity, no extreme bias.

## Managerial Insights

- Actively manage academic pressure.
- Foster study satisfaction environments.
- Early intervention for students with lower CGPA.

# QS World Rankings Dataset (Linear Regression)

**Objective: Predict a university's Overall Score using key performance indicators.**

Primary Research Question: "*Which university performance metrics (academic reputation, employer reputation, sustainability, etc.) best predict overall QS rankings?*"

Source: Kaggle

## Key Variables

Academic Reputation

Employer Reputation

Sustainability Score

Faculty Student Ratio

Citations per Faculty

International Students

## EDA Highlights (Tableau)

Scatterplot: Strong positive link between Academic Reputation and Employer Reputation.

Bar Chart: Oceania universities excel in International Student scores.

Bubble Charts: Regional strengths in internationalization visible.

Additional Insight: Majority universities are public institutions.

# Dashboard 1

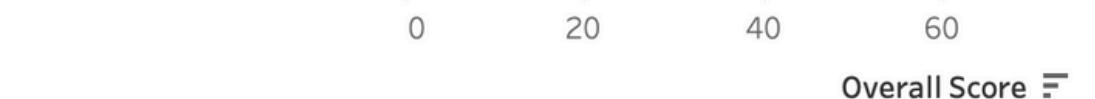
LEADING REGIONS  
AND UNIVERSITIES  
BASED ON OVERALL  
SCORE,  
SUSTAINABILITY  
SCORE, ACADEMIC  
REPUTATION,  
EMPLOYER  
REPUTATION

## Leading Countries & Institutions in QS 2025

### Top 20 Universities by Overall Score

#### Institution Name

Stanford University	96.10
ETH Zurich - Swiss Federal..	93.90
National University of Sin..	93.70
UCL	91.60
California Institute of Tec..	90.90
University of Pennsylvania	90.30
University of California, B..	90.10
The University of Melbour..	88.90
Peking University	88.50
Nanyang Technological Un..	88.40
Cornell University	87.90
University of Michigan, Ann..	87.80
Harvard University	87.70
University of Oxford	87.50
Massachusetts Institute of T..	87.40
University College London	87.30
University of Cambridge	87.20
University of Texas at Austin	87.10
University of Illinois Urbana..	87.00
University of Washington	86.90



#### Region

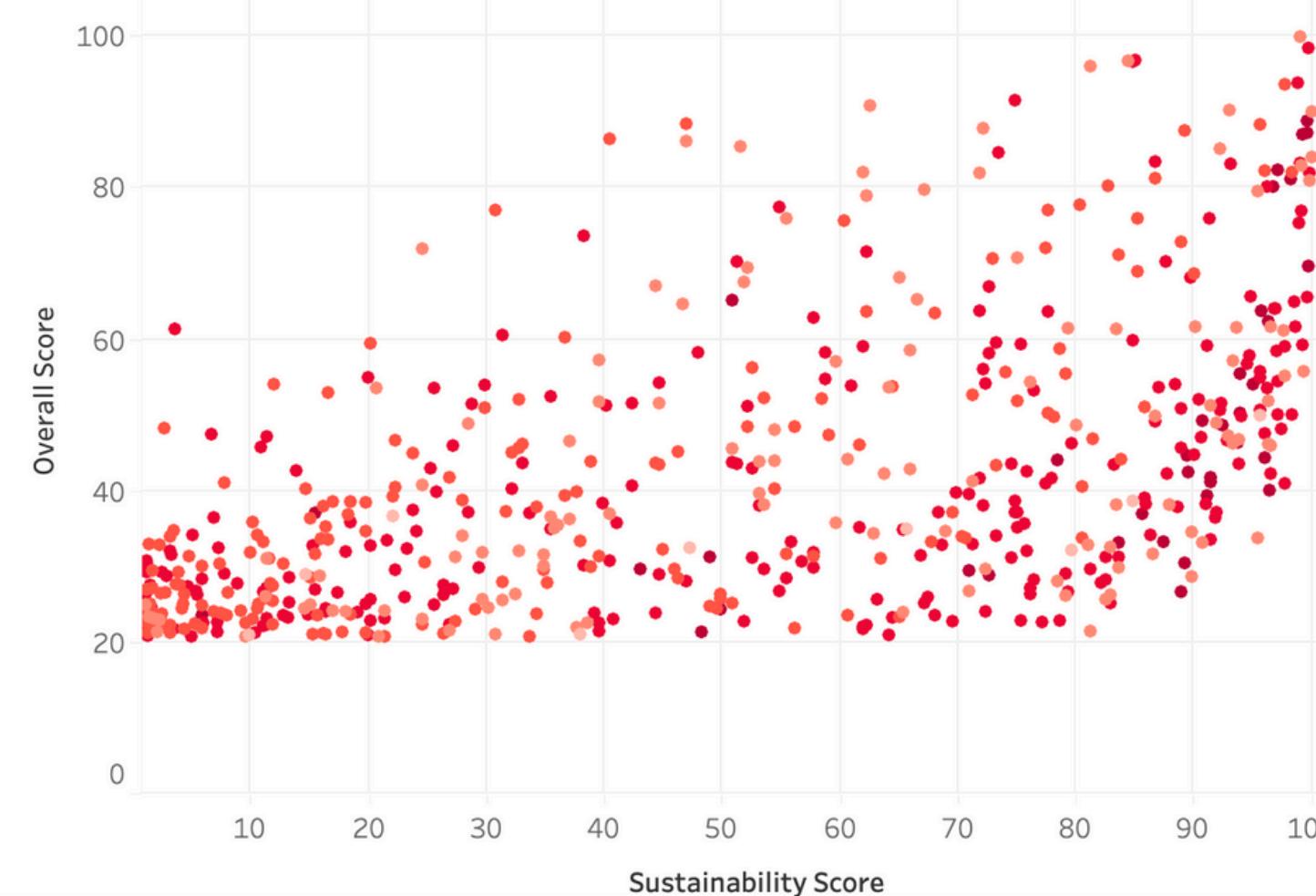
Africa  
Americas

Asia  
Europe  
Oceania

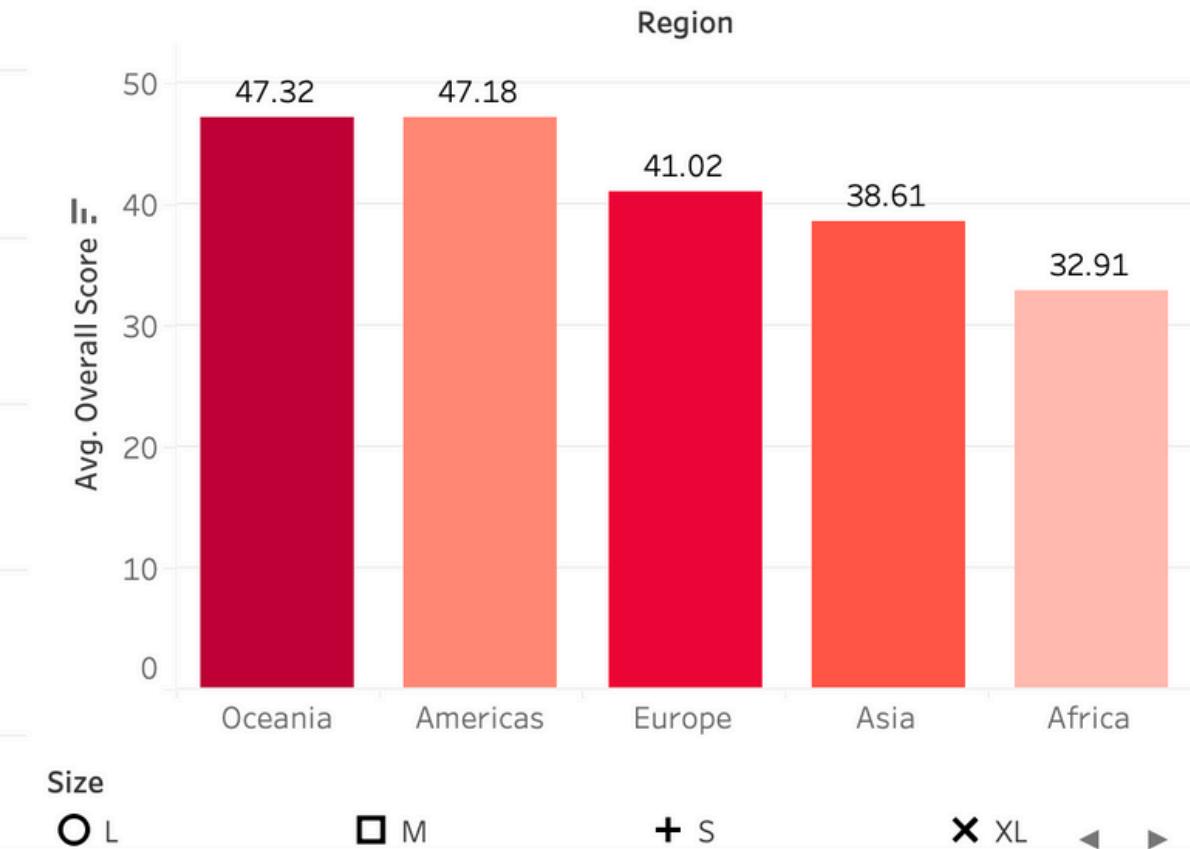
### Academic vs Employer Reputation



### Sustainability Score vs Overall Score



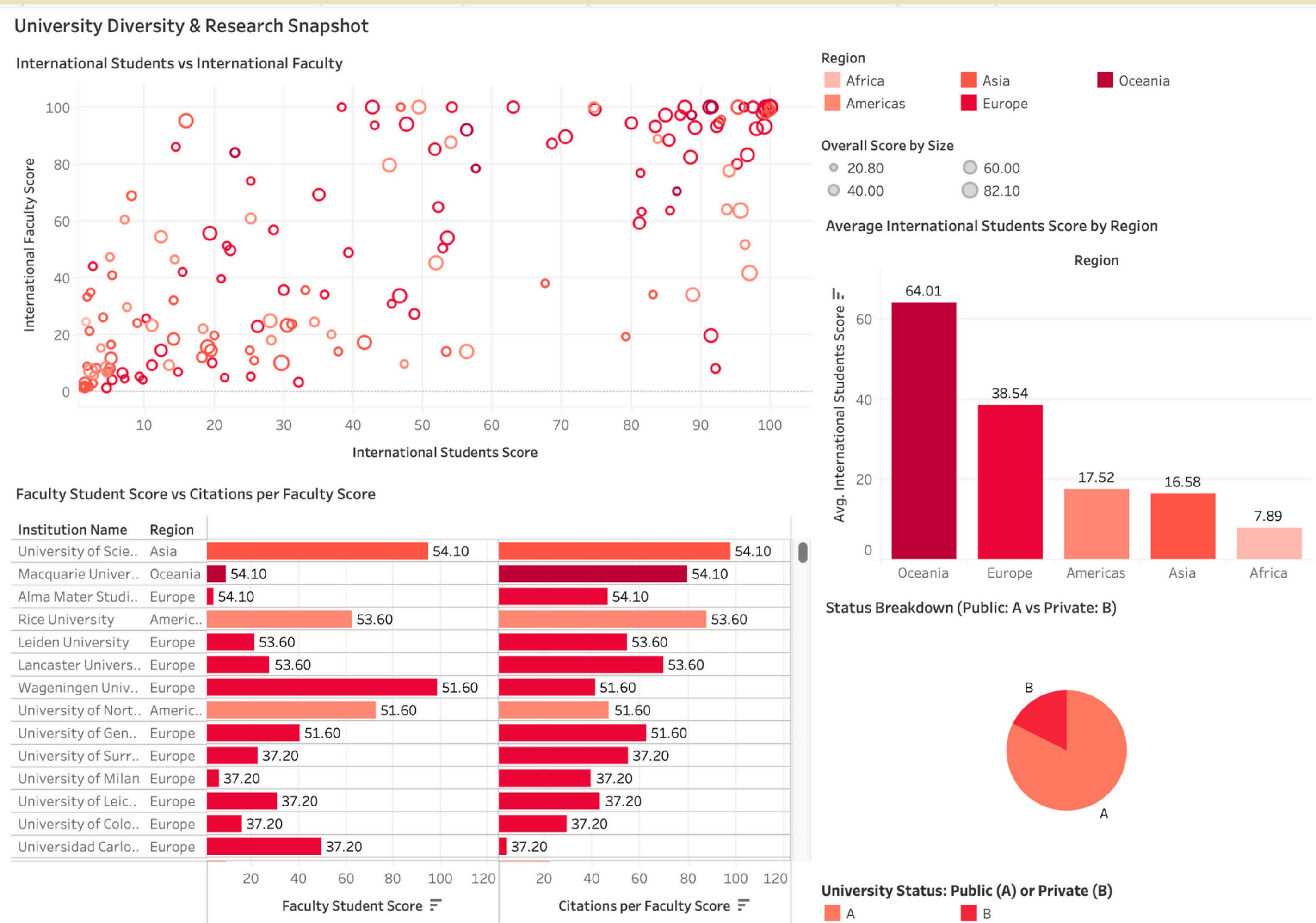
### Average Overall Score by Region



# Dashboard 2

DIVERSITY IN UNIVERSITIES: WHICH COUNTRIES ATTRACT THE MOST INTERNATIONAL STUDENTS?

STATUS BREAKDOWN: PUBLIC AND PRIVATE



# Model Building: RStudio

## Method: Multiple Linear Regression

Final Model (post Stepwise AIC selection):

Academic Reputation  
Employer Reputation  
Sustainability Score

## Performance:

R-squared: 85.4%  
Adjusted R-squared: 85.2%

## Diagnostics

Residual vs Fitted: No patterns – Linearity assumption met.  
Normal Q-Q Plot: Residuals follow normality with minor edge deviations.  
Scale-Location Plot: Constant variance – Homoscedasticity assumed.  
Residuals vs Leverage: No influential outliers.  
VIF Values: All <5 – no multicollinearity.

## Managerial Insights

Boost Academic and Employer Reputation.  
Strengthen Sustainability initiatives.  
Improved scores = more funding, better faculty, global visibility.

# EXECUTIVE SUMMARY

Extensive Exploratory Data Analysis (EDA) using Tableau revealed **key patterns** in both datasets. Regression models (Linear and Logistic) successfully identified significant predictors:

*QS World Rankings: Academic Reputation, Employer Reputation, Sustainability Score.*  
*Student Depression: Academic Pressure, Study Satisfaction, CGPA.*

All key model assumptions were checked and satisfied, including linearity, homoscedasticity, and multicollinearity.

Model performance was strong:  
Linear Regression: R-squared of **85.4%** (QS dataset).  
Logistic Regression: Classification accuracy of approximately **75–80%** (Depression dataset).

Actionable insights were developed:  
Strategic recommendations for improving university rankings.  
Early intervention strategies for supporting student mental health.

# Skill Set

## Data Visualization & Storytelling

### Primary Tools

- Tableau (dashboards, calculated fields, interactive filters)
- Power BI (DAX, custom visuals, report design)
- Excel (PivotCharts, conditional formatting, dynamic dashboards)

### Programming & Libraries

- Python: matplotlib, seaborn, plotly
- R: ggplot2, tidyverse visualizations

### Design & Prototyping

- Figma (dashboard mockups, UI wireframes)
- Adobe Illustrator (infographics, custom charts)
- Canva (presentation-ready visuals)

### Core Competencies

- Interactive dashboards for stakeholder decision-making
- KPI tracking & performance monitoring
- Exploratory data analysis (EDA) visualization
- Storytelling through data (reports, presentations, pitch decks)

# Contact Me

Phone

+91 9823919333

College Email

kamya.sarda@flame.edu.in

Personal Email

kamya.hemant@gmail.com

