

GRADUATION BY ZONE

United Arab Emirates

EDA Team 2

Team Lead: Mike 80s

Team Members:

Amulya kulkarni

Jane Ndambuki

Mary Aleta White

Michael Fabián Ramos Hernández

Sahar Nikoo

Overview and Introduction

Almost universally, student academic achievement is considered critical to their educational and career success. One of the most common measures of student achievement is their ability to pass an exam, a course, or an entire academic program (e.g., college graduation). There are other measures of student success such as persistence and retention in their program of study. In this study, we used a data set to predict student pass rates from secondary school. The purpose of this study is to use machine learning algorithms to predict student passing rates within the seven United Arab Emirate states. The findings can improve an understanding of which factors can best predict student pass rates. This information may be of use to Ministries of Education who are interested in increasing students' academic performance and strengthening educational achievement within their region.

The data set used in this study is graduation data gathered by the Ministry of Education beginning in 2015 and ending in 2018. As described in more detail below, the data file provides key information on students who graduated or who failed to graduate after the 12th grade (secondary education or high school in the U.S./Europe and primary school in the UAE).

Exploratory data analysis

The initial work of the Exploratory Data Analysis Team (Team 2) was to review the selected data set and begin to explore its features using basic descriptive statistics, Visualizations in Plotly, Heatmaps in Plotly, and feature engineering.

Dataset properties

The link to the raw data set is provided here: <http://data.government.ae/dataset/graduation-rate-by-zone-school-type-school-stage-gender-nationality-group>

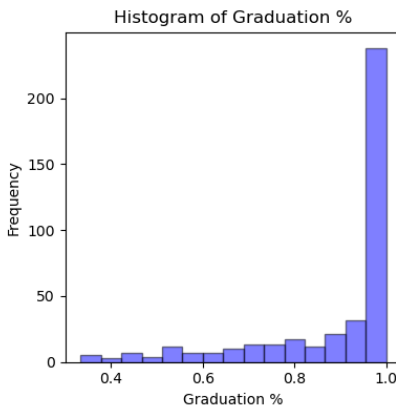
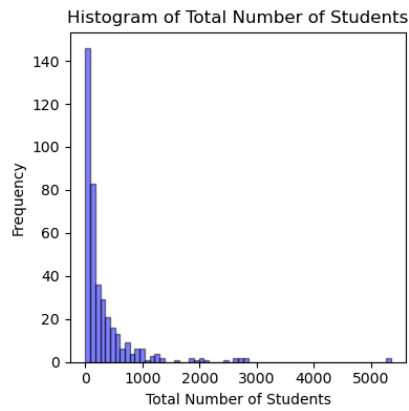
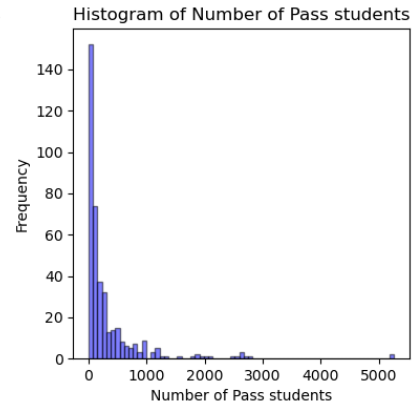
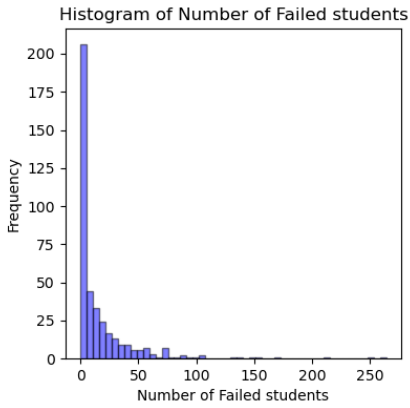
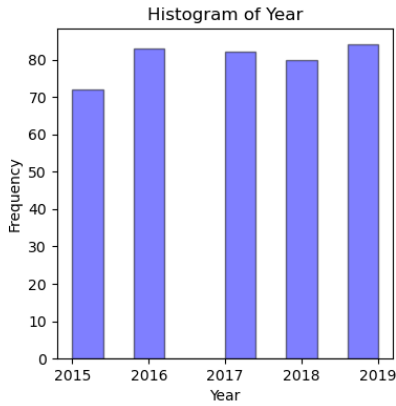
The graduation data was gathered for the first time in 2015 by the Ministry of Education. The data set represents students who graduated by the Emirate, school type, gender, and nationality (local or Expat). The students were those completing the 12th grade for academic years 2015-2018.

Based on notes between team members in the Slack channel, the EDA team first noticed that the dataset had a number of null values including a blank/empty column that was dropped. Similarly, for one variable, School stage, all schools were listed as Cycle 3, so that item was also dropped. To address the null values, the team met and identified one Emirate with missing coordinates data. They filled them in by locating the Emirate coordinates, latitude and longitude so the variable could be used in analysis.

The following section provides an overview of the original data set variables:

#	Column	Non-Null Count
0	Year	401 non-null
1	Zone_EN	401 non-null
2	Coordinates	333 non-null
3	School_Type_EN	401 non-null
4	School_Stage_EN	401 non-null
5	Gender_EN	401 non-null
6	Unnamed: 6	0 non-null
7	LocalFlag_EN	401 non-null
8	Number of Failed students	401 non-null
9	Number of Pass students	401 non-null
10	Total Number of Students	401 non-null
11	Graduation %	401 non-null

Descriptive statistics for each variable is provided here including frequency histograms for five key variables.



Value counts for Year:

2019	84
2016	83
2017	82
2018	80
2015	72

Value counts for Zone_EN:

Ajman	73
DUBAI	68
Fujairah	66
Sharjah	66
Ras AlKhaima	65
Umm AlQuwain	63

Value counts for Coordinates:

25.244000,55.266140	73
25.133300,56.250000	66
25.322327,55.513641	66
25.800694,55.976200	65
25.520482,55.713391	63

Value counts for School_Type_EN:

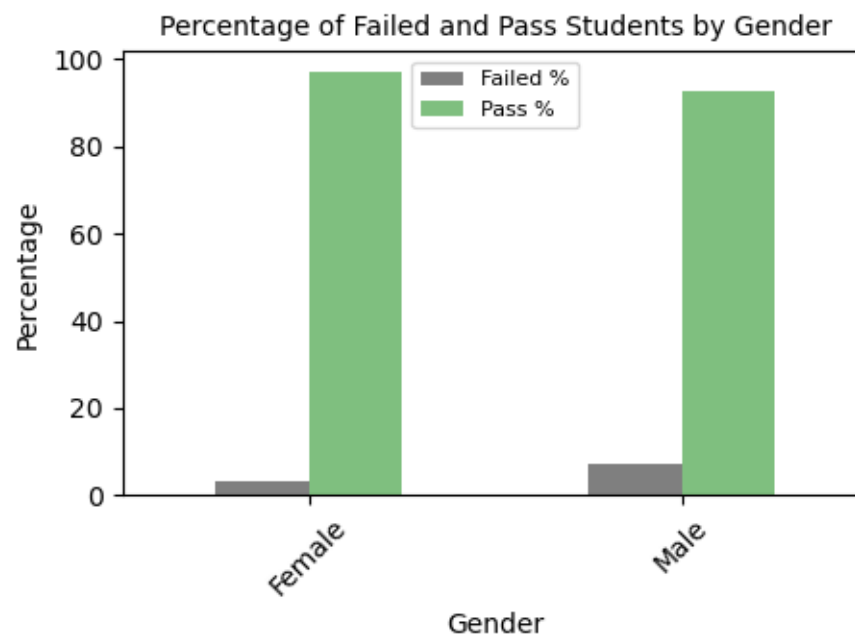
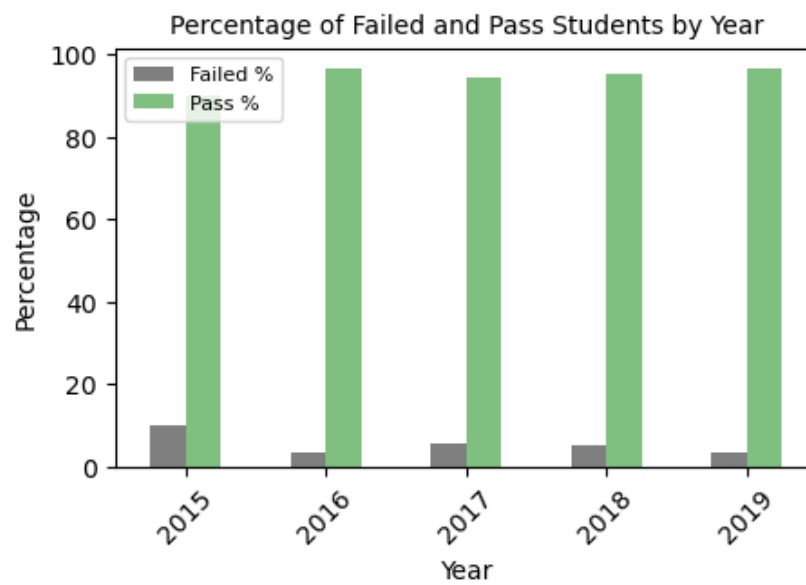
Public Schools	120
Continuing Education Centers	120
Private Schools	116
ACTVET	45

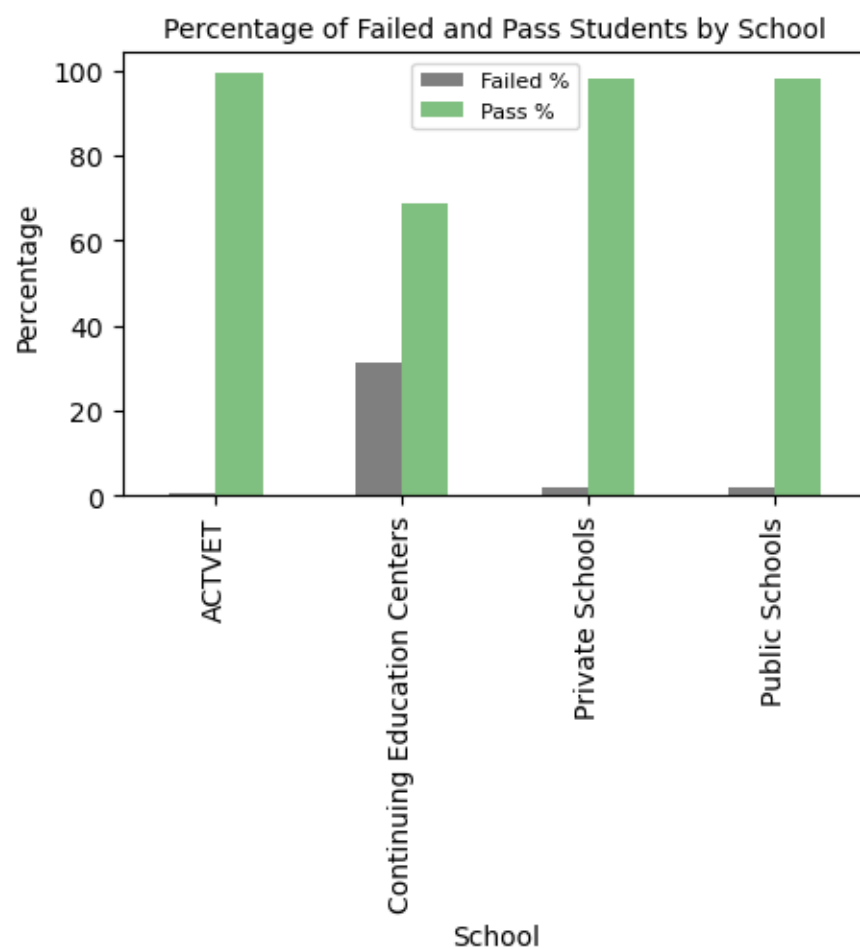
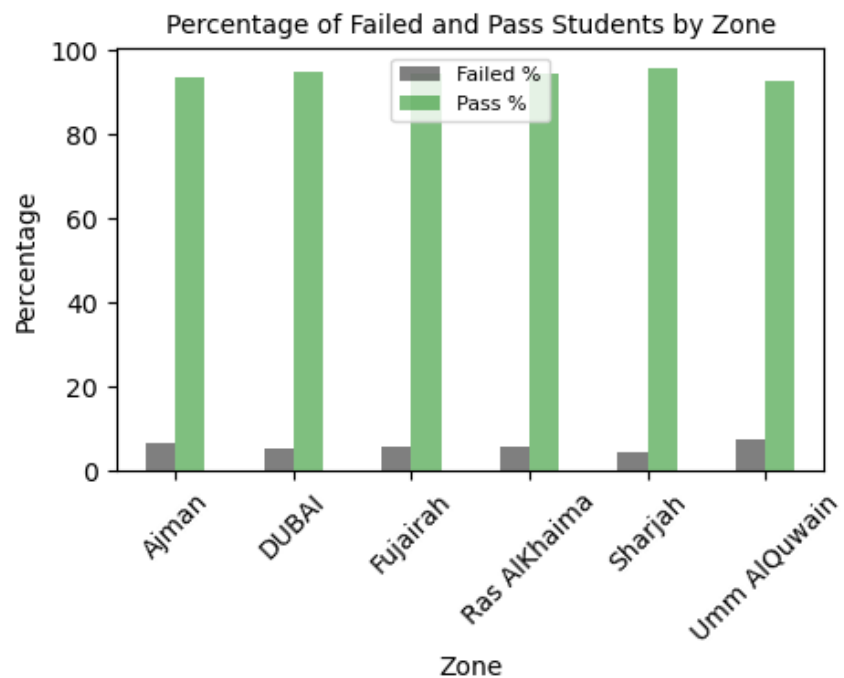
Value counts for School_Stage_EN:
Cycle 3 401

Value counts for Gender_EN:
Male 212
Female 189

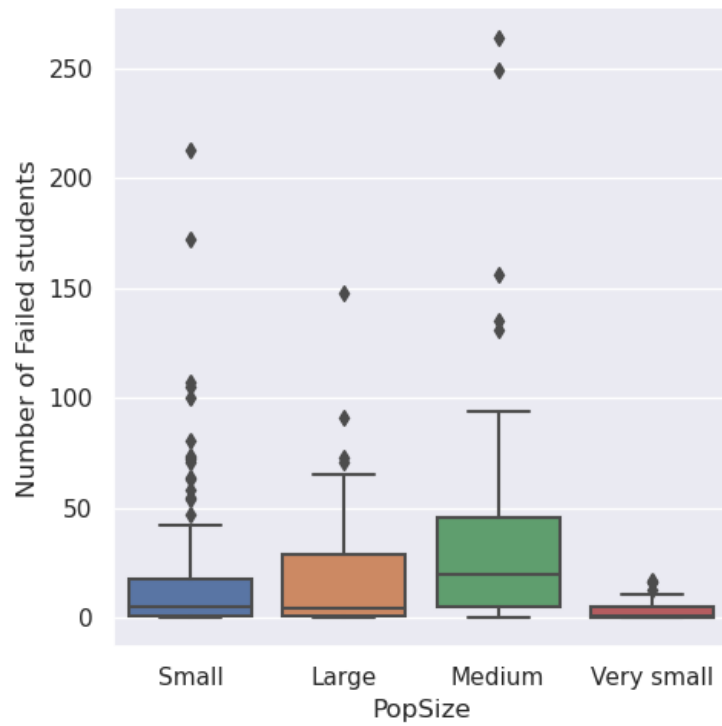
Value counts for LocalFlag_EN:
Local 206
Expat 195

The charts below further describe the data set by displaying student fail/pass rates by variables such as year, gender, Emirate, school type, and the relative population size of each Emirate.

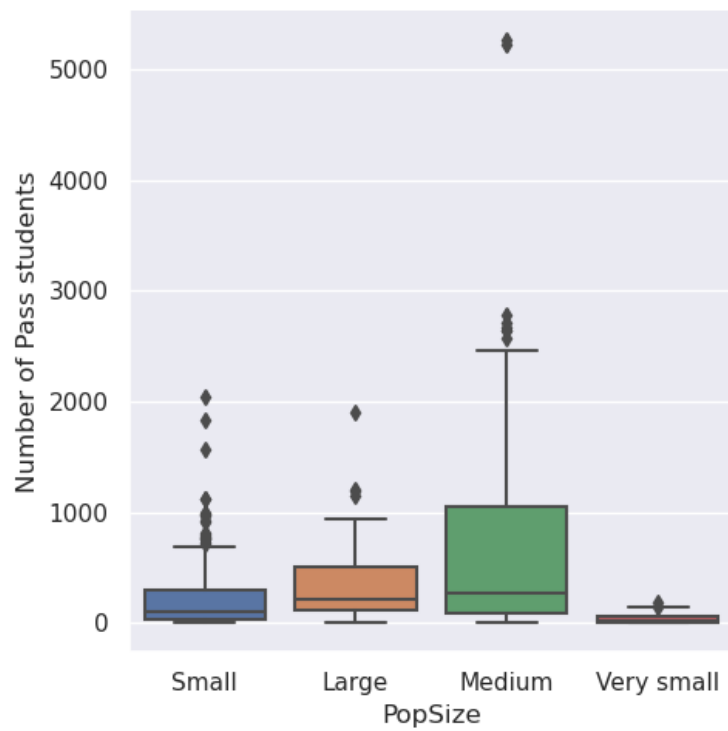




Plot 1: Failed attempts by Emirate size

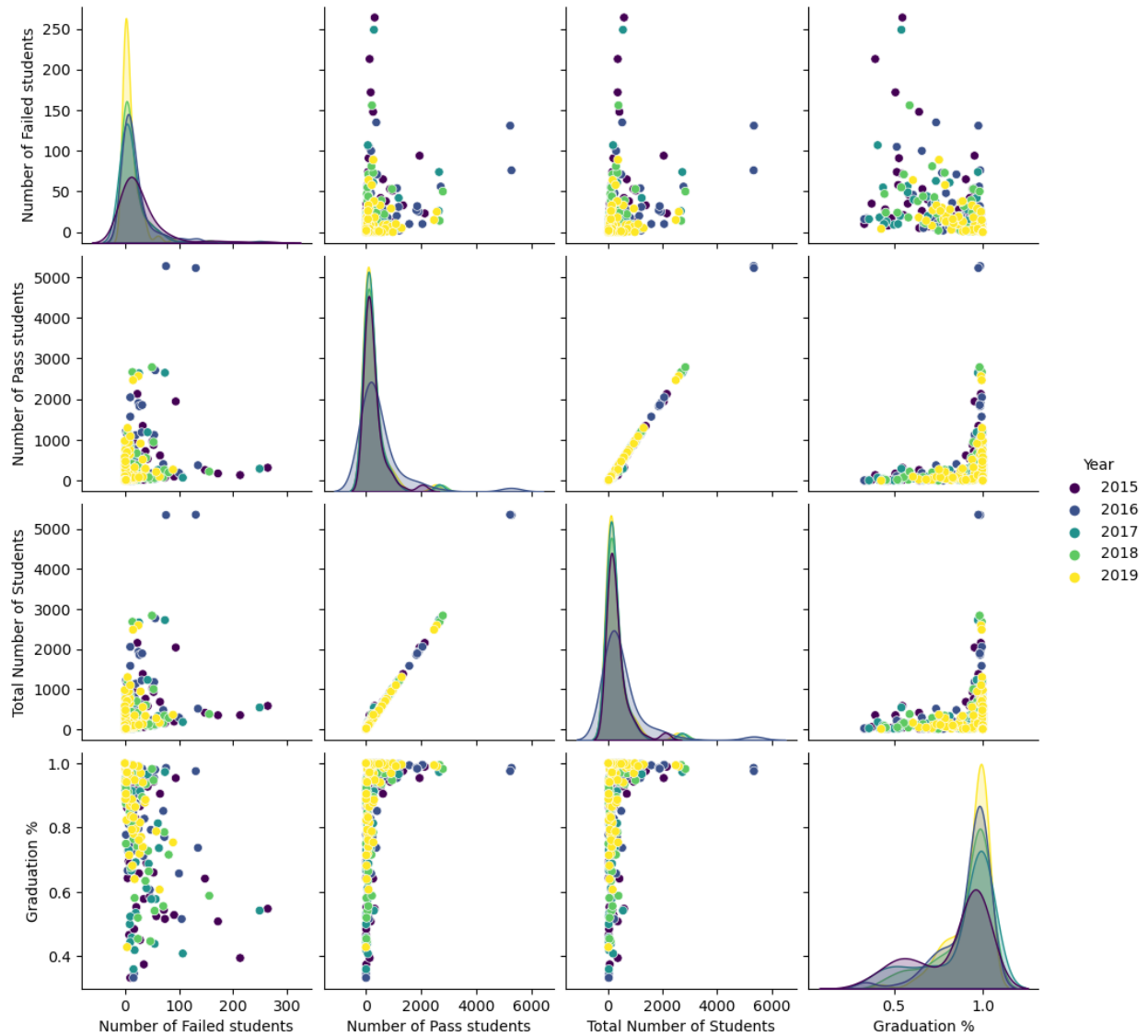


Plot 2: Pass attempts by Emirate size

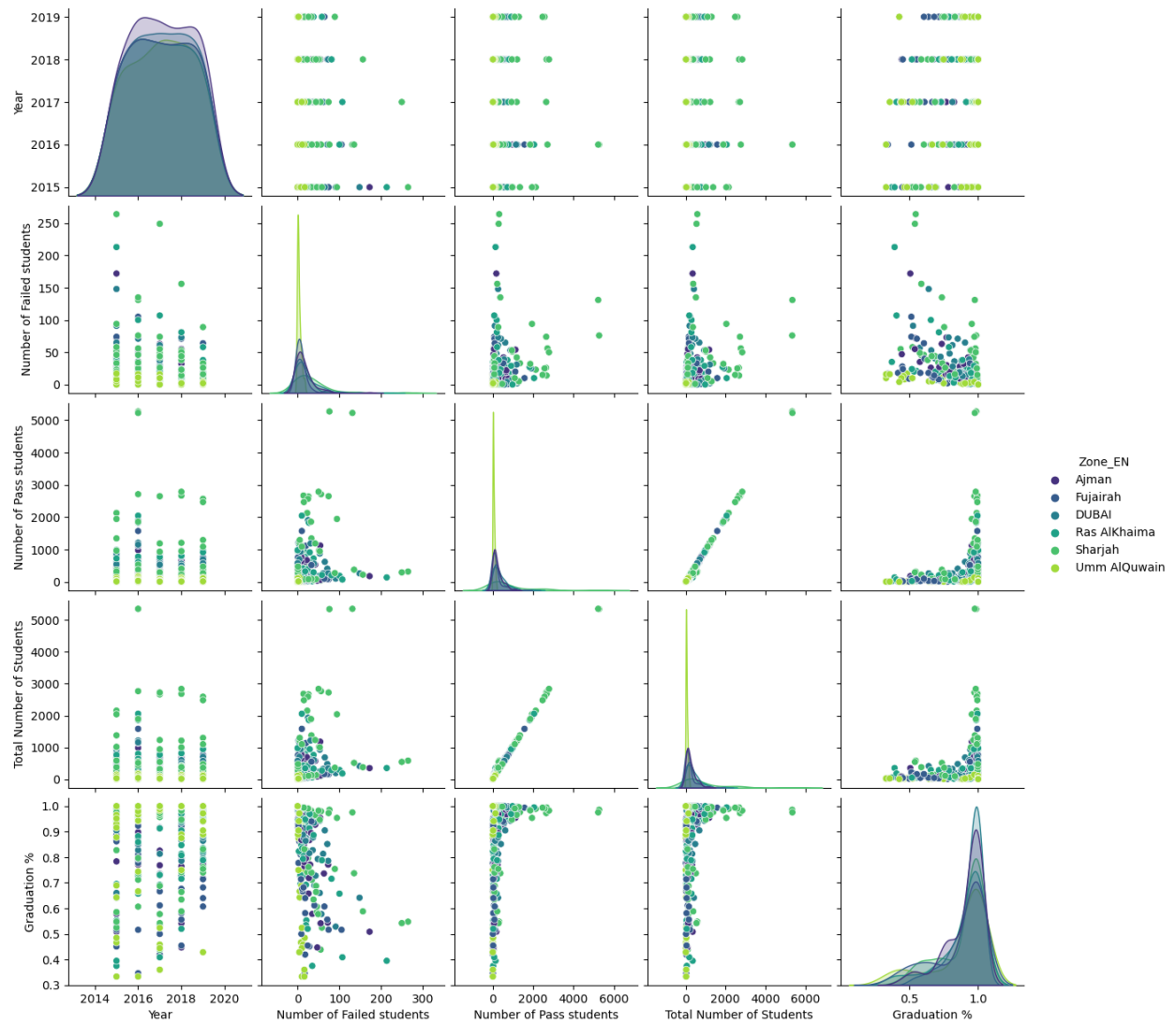


In the next phase of exploratory data analysis, researchers created pair plots (let's decide on the most relevant?) as shown below:

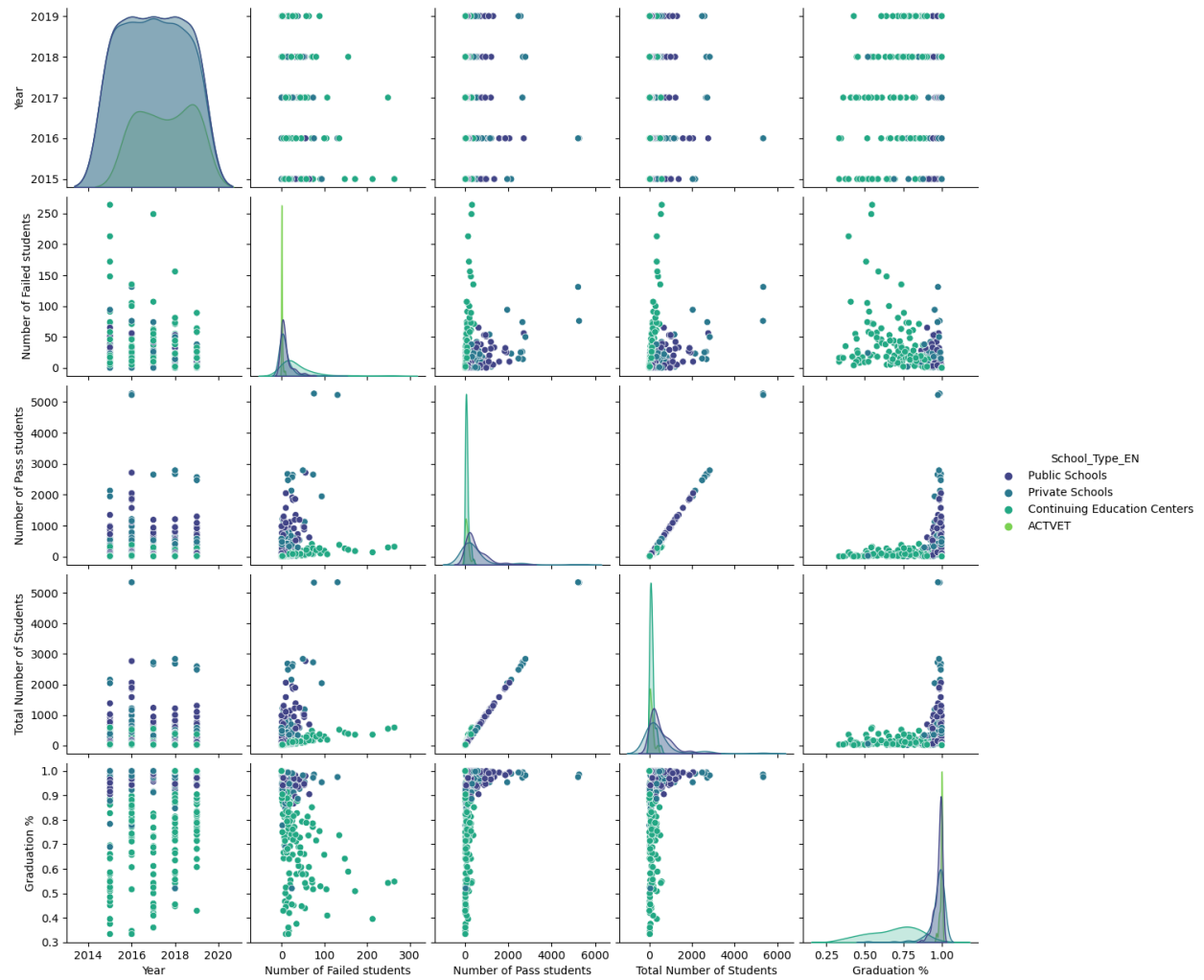
1. Pair plot by year



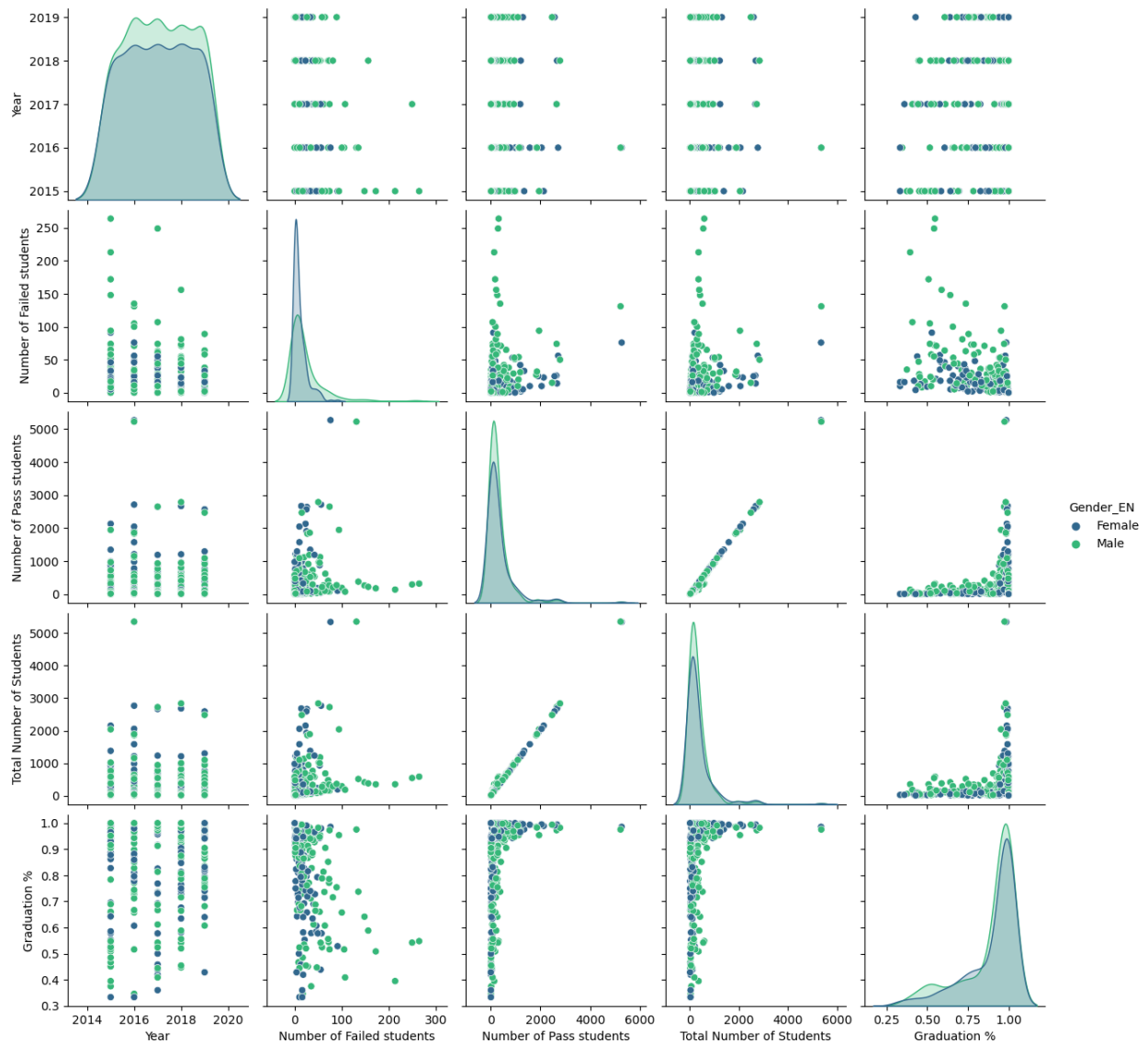
2. Pair plot by Emirate



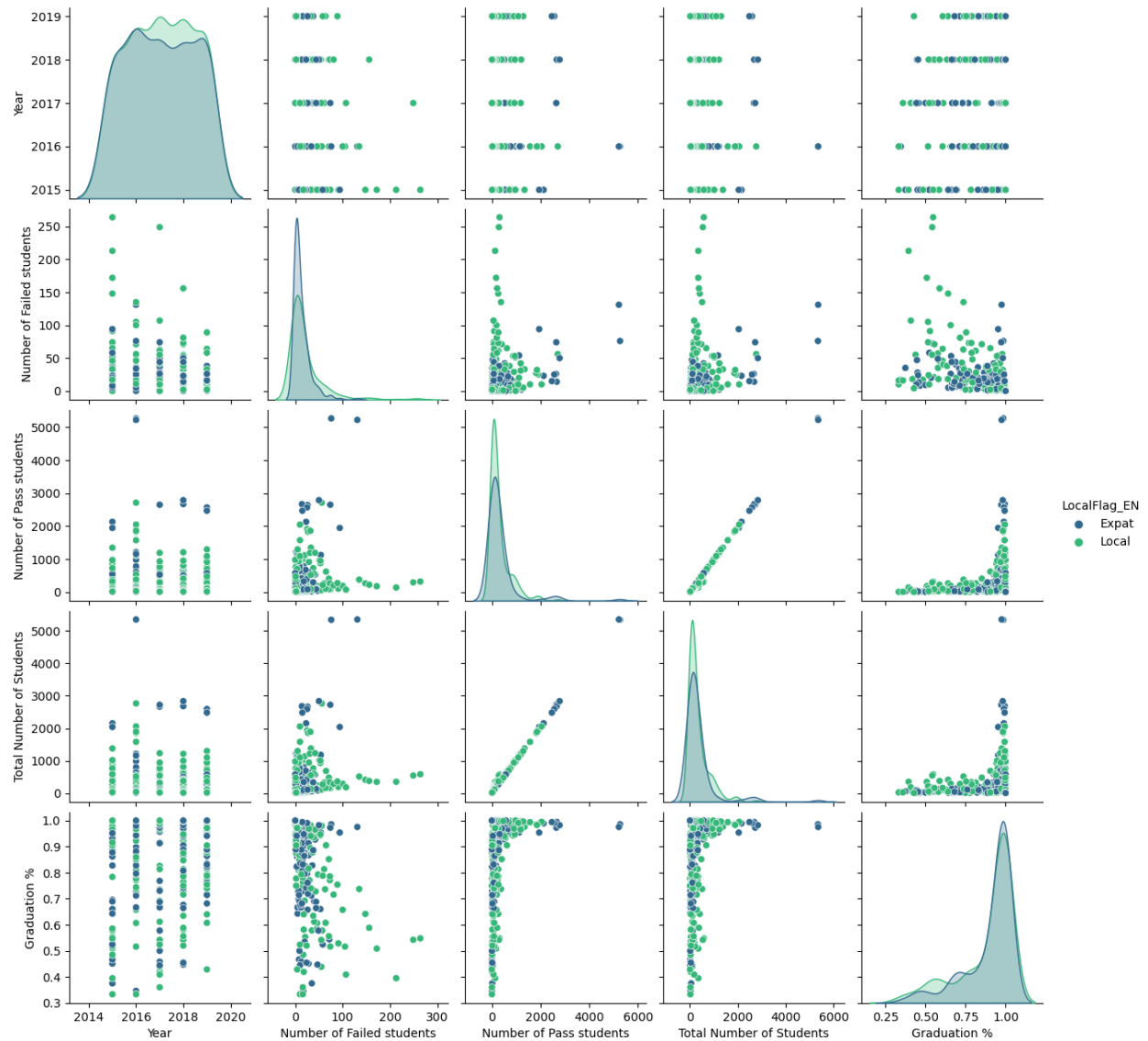
3. Pair plot by School Type



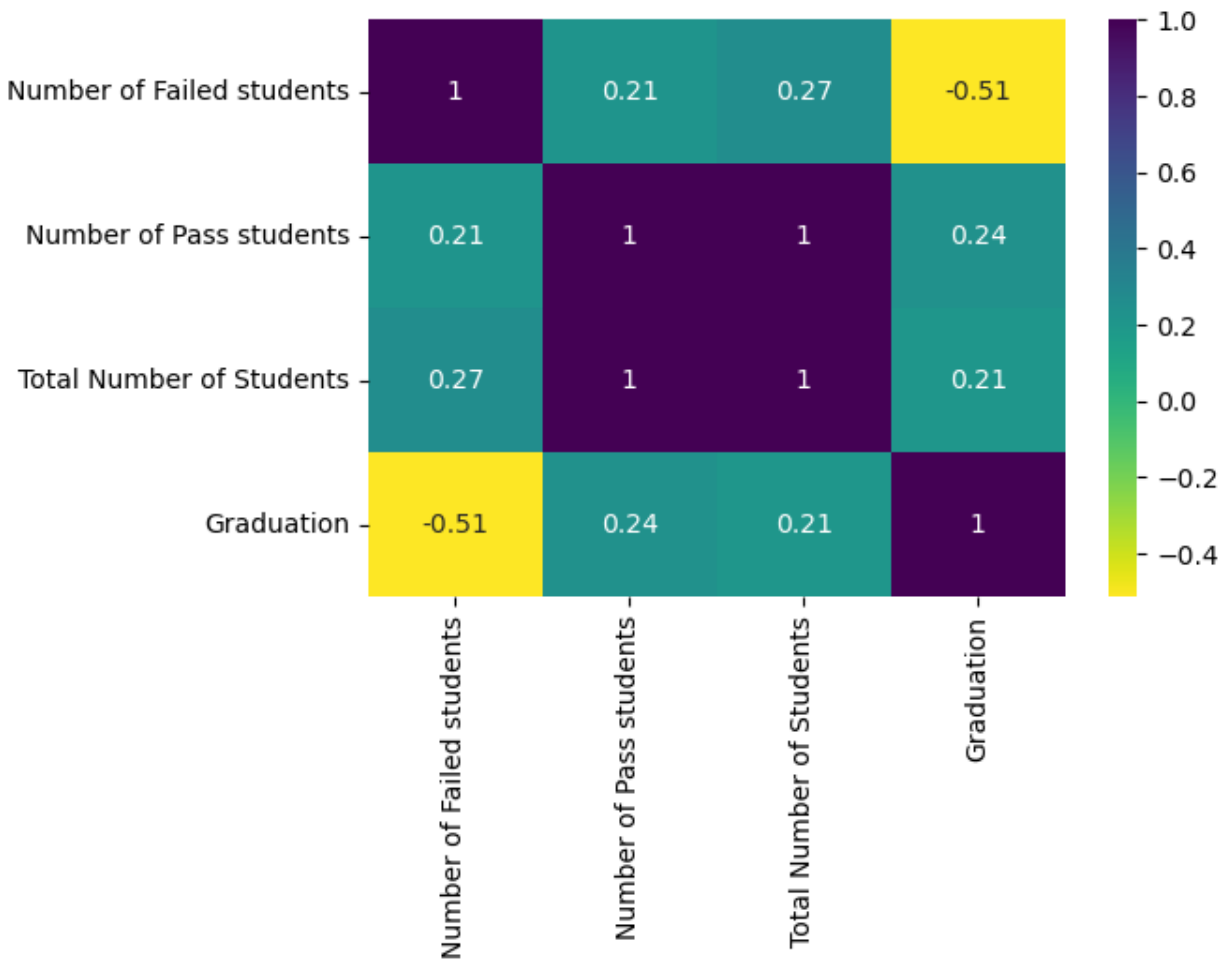
4. Pair plot by Gender



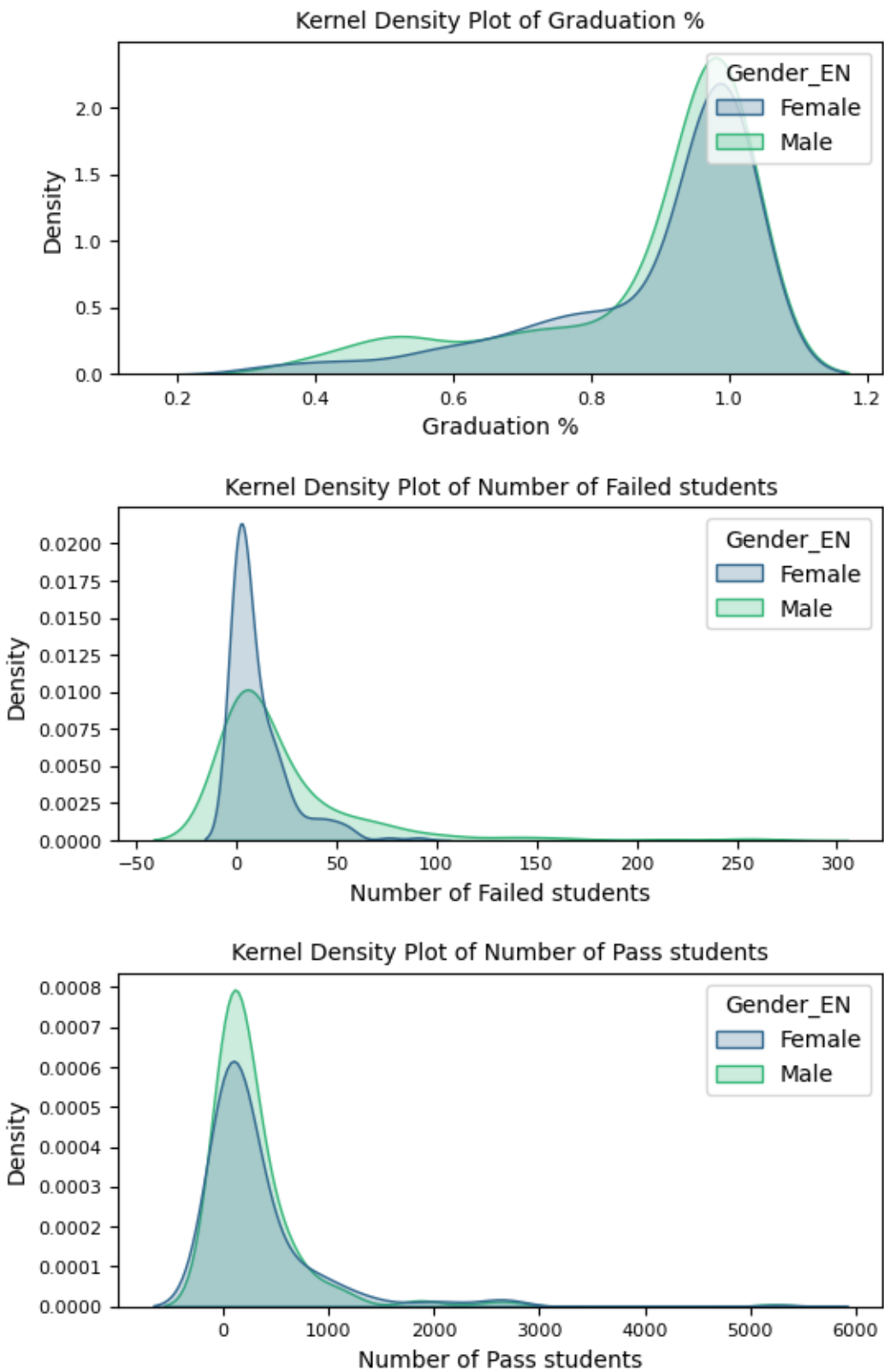
5. Pair plot by Nationality



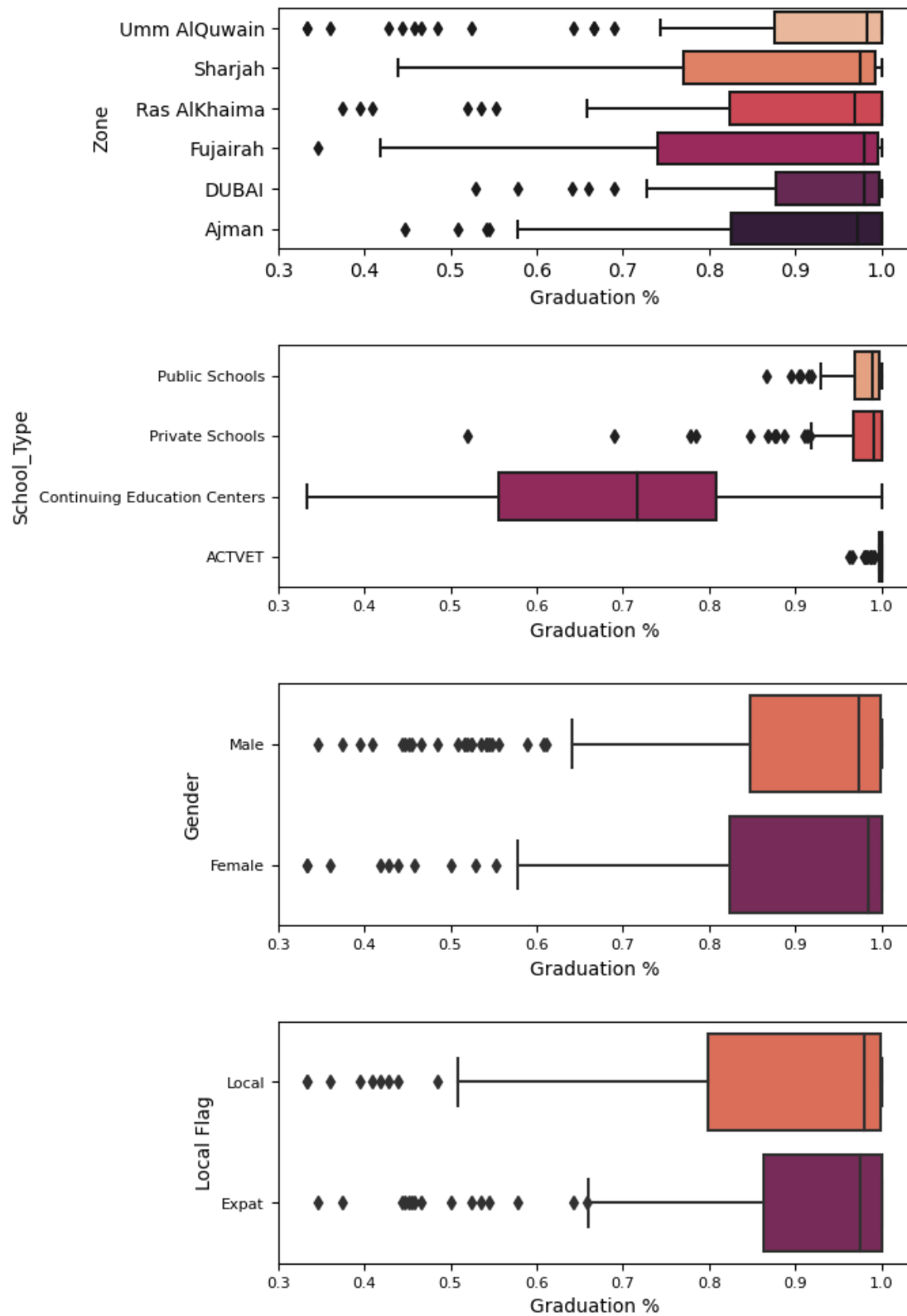
Correlation matrix



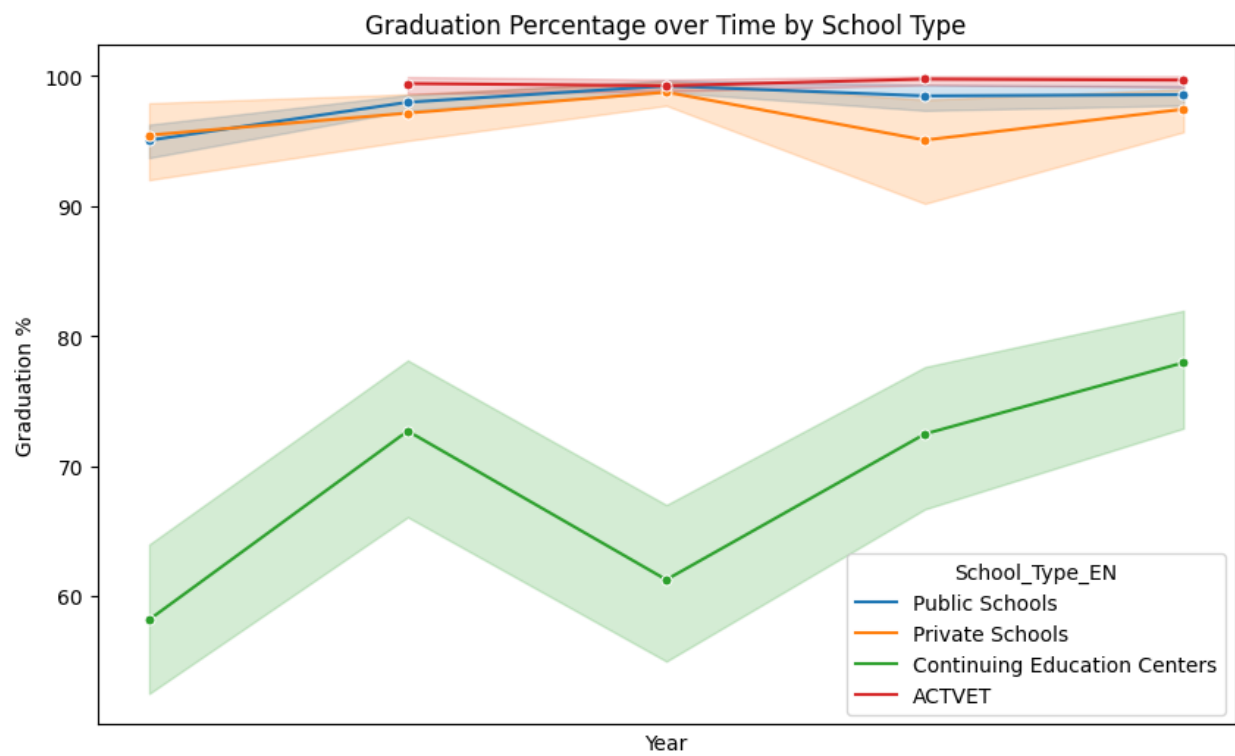
Kernal density plots

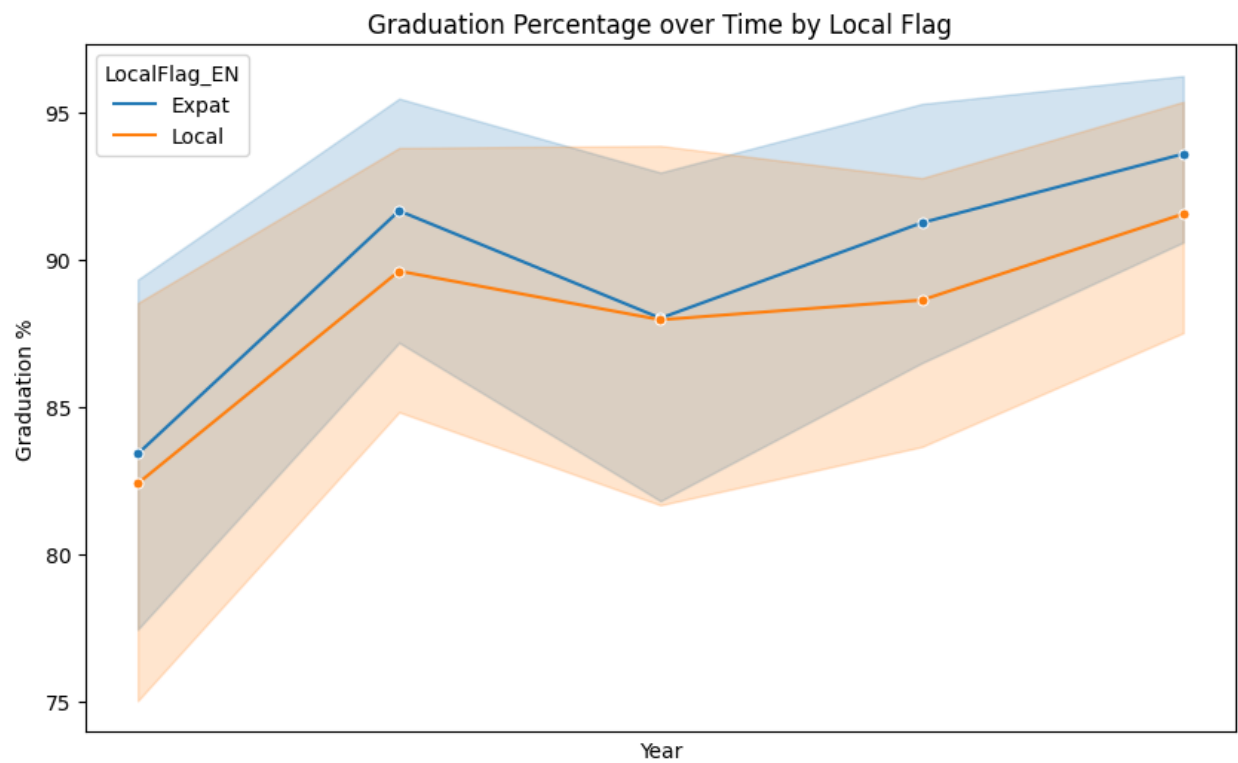
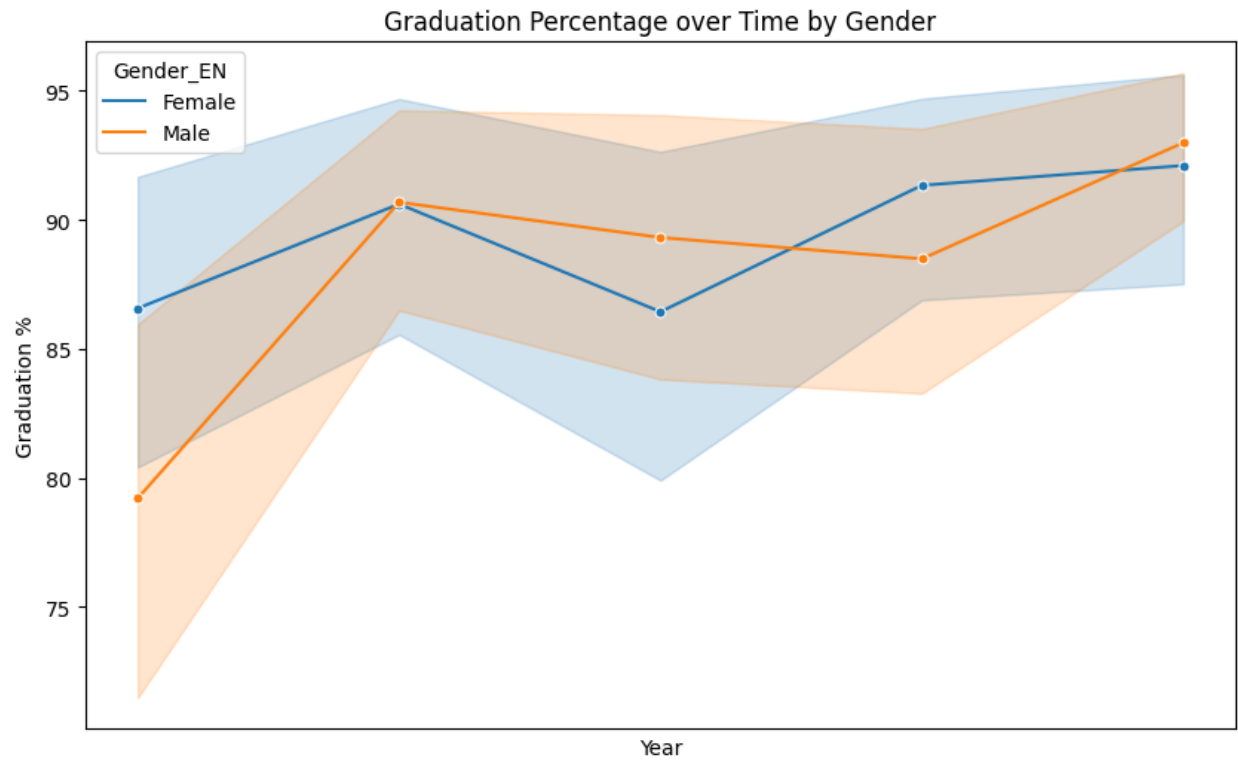


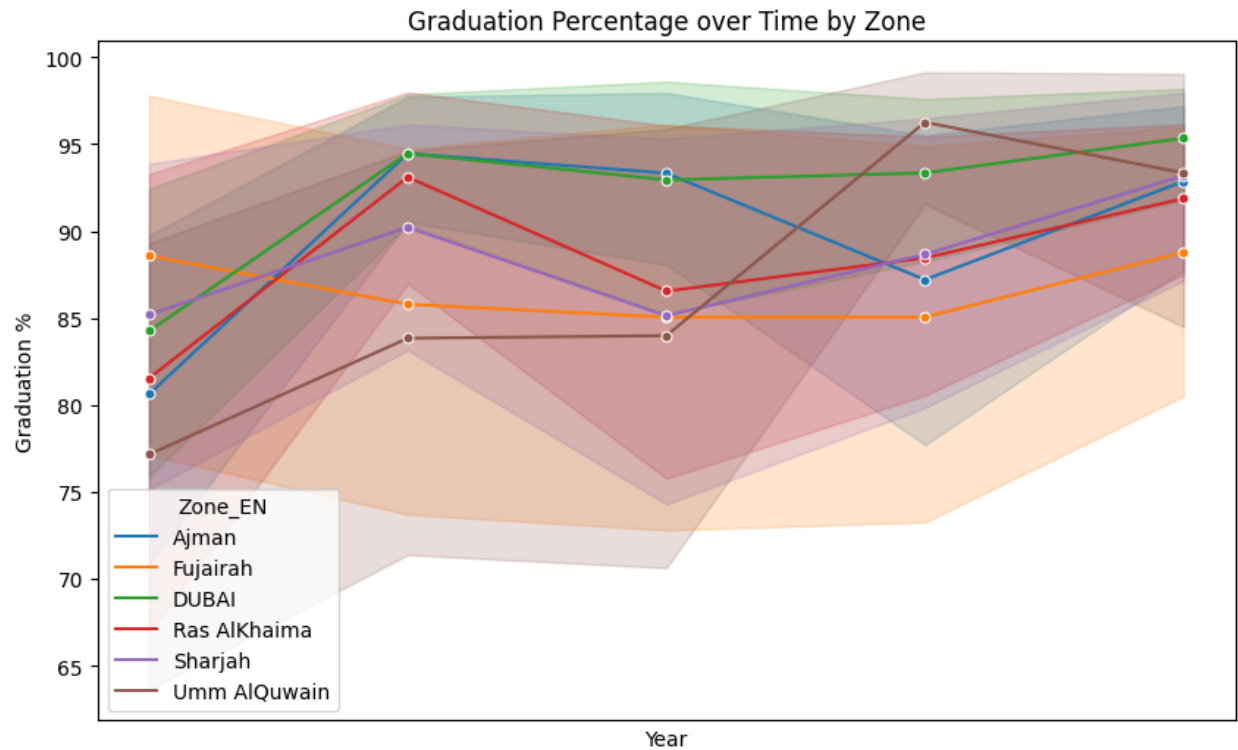
Box plots for categorical variables



Graduate Rates over time by School type (Amulya)







Bar Graphs



Concludes

After cleaning the data, generating several graphs and finding some trends, we define that our objective column is graduation.

After carrying out the analysis of variance and correlation between the variables, we can affirm that:

By definition

Number of student is a significant variable

Pass student is a significant variable

Failed is a significant variable

By testing

Year is a significant variable.

School_type is a significant variable.

Gender is not a significant variable.

LocalFlag is not a significant variable

Zone is not a significant variable

Since the zone is not a significant variable, we perform a spatial analysis and deduce that:

The closer to Dubai the graduation rate increases and the further from Dubai the graduation rate decreases

UNITED ARAB EMIRATES

