Department of Computer Science
FAST National University of Computer and Emerging Sciences
Karachi Campus

Predictive Analytics on the Academic Record of NUCES

Submitted by :

Obaid Ur Rehman
17k-3848

Areeka Ajaz
17k-3913

Tooba Shahid
17k-3731

Submitted On :

25th June 2021

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science

Supervisor :

Dr Jawwad Ahmed Shamsi

ACKNOWLEDGEMENT

We are Thankful to Allah and His blessings.

We would like to express our utmost gratitude to our supervisor Dr. Jawwad A. Shamsi, for his enthusiasm, patience, insightful comments, helpful information, practical advice, and unceasing ideas that have helped us tremendously at all times especially during research and development of this project. His immense knowledge, profound experience, and professional expertise has enabled us to complete this project successfully. Without his support and guidance, this project would not have been possible.

We are also thankful to our FYP Jury and FYP Committee for their guidance at each evaluation.

ABSTRACT

Predictive Analytics is the process of using past data to make future predictions . The past data is used to capture important trends with the help of a mathematical model and the model is then used to make predictions on current data .Predictive analytics has become an influencing factor in improving educational experiences for students. The result of predictive analytics on academic record plays a big role in a way to achieve the highest level of quality of education. This analytics can not only be used to better understand student performance but also to boost graduation rates. Moreover , the predictive model may also help to identify the students who are subject to low performance at an early stage and do the necessary intervention. Hence , early student performance prediction can help universities to take appropriate actions on time to improve the success rates of students . We aim to perform predictive analytics on the academic record of NUCES . Our project will use the past academic record of NUCES to make insights and find out correlations between different attributes and then build a predictive model , based on the statistical inference , to make predictions on the current data. Finally we aim to integrate our analytics with a system , developing a fully functional web portal .

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

Each year a number of students take admission in FAST NUCES . Being a national university it has campuses spread in FIVE major cities of Pakistan . Students from all over Pakistan , belonging to different cities , take admission in different campuses at FAST NUCES. These students belong to different educational backgrounds , different families , have different capabilities and have different past educational records. Their academic performance throughout their university life is a reflection of different factors , not only but including their educational background , their previous academic records , the region/district from where they belong etc . By the use of statistical analysis , this project aimed to answer a series of questions about how these factors are related to the performance of students at FAST throughout their educational period. Other than this the project aimed to find out how students in FIVE different campuses of FAST perform and make a campus wise comparison. The project also aims to find out how a degree program affects a students performance and what is the role of different course domains and then courses offered under them. The project finally aims to use this information to produce a model to predict the student's CGPAs and categorize them into different groups. Through modelling , the project also aims to find out the best fit model for the problems related to predictive analysis on academic data . Last but not the least , the project concludes with building a fully functional system for predictive analytics on academic data.

CHAPTER 2

LITERATURE REVIEW

Predictive analytics has become an influencing factor in improving educational experiences for students. The result of predictive analytics on academic record plays a big role in a way to achieve the highest level of quality of education. This analytics can not only be used to better understand student performance but also to boost graduation rates. Moreover , the predictive model may also help to identify the students who are subject to low performance at an early stage and do the necessary intervention. Hence , early

Student performance prediction can help universities to take appropriate actions on time to improve the success rates of students .

A standard predictive analytics process starts by integrating raw data – from different data sources. This data becomes the basis of the analytics , as this data can be utilized for discovering unknown patterns and trends as well as hidden relationships. However the data in its original form is usually not ready for analysis and modeling. Since the data is usually formed as a combination of different tables , the data contains duplications , missing values and inconsistencies. It is important to know how to handle them without compromising the quality of the prediction. Therefore the data has to go through an initial preparation (cleaning) , before it can be further utilized. All things considered, this cannot be done by a general procedure, and several methods need to be considered within the context of the problem. The main approaches of cleaning data involve listwise deletion and imputation.

Once the data is cleaned , preliminary statistical analysis, especially through visualization, is done which allows us to better understand the data. This helps in identifying outliers and imbalance in the data which must be removed for better accuracy of the analysis.

After the preliminary statistical analysis the data preprocessing step begins. In this step , the data undergoes transformation of which the most commonly used methods are normalization and encoding. Then to remove imbalance from the data set either over sampling or under sampling is done. Now that the data has been cleaned and transformed it is ready to be used for finding patterns and trends.

To discover different patterns that can improve students' performance, many studies have been conducted. Especially during the last few years lots of research has been carried out to predict students' academic performance. The research begins with identifying the important factors (feature selection) that affect the students' academic performance. Feature selection, an important strategy to be followed , aims to choose a subset of attributes from the input data. Feature selection enables reduced computation time, improved prediction performance while allowing a better understanding of the data. For our problem , different researchers have identified different factors that affect academic performance .

Abeer Badr El Din Ahmed et. al. , in his study , used the course of the student, mid-term marks, Lab test grade, assignment, attendance, homework, student participation. Another research was carried out by Fadhilah Ahmad and Azwa Abdul Aziz in which they used nine parameters like gender, race and hometown, GPA, family income, university entry mode, and

grades in related courses. Mohammed M. Abu Tair and Alaa M. El-Halees in their study tried to extract some useful information from student's data of Science and Technology College – Khan Younis. They initially selected different attributes like Gender, date of Birth, Place of Birth, Speciality, Enrollment year, Graduation year, City, Location, Address, Telephone number, HSSC Marks, SSC school type, HSSC obtained the place, HSSC year, College CGPA for analysis. But after preprocessing the data they found that attributes like Gender, Speciality, City, HSSC Marks, SSC school type, College CGPA are most significant. . Jyoti Bansode for predicting student academic performance collected data from Shah and Anchor Kutchhi Polytechnic, Chembur, Mumbai. They considered student attributes like parent's education, parent' s occupation, category, SSC board, admission type, SSC medium, SSC class, first-semester result, second-semester , third-semester, fourth-semester, the fifth-semester and sixth-semester result as most important attributes. Maria Koutina and Katia Lida Kermanidis tried to find out the best techniques for predicting the final grade of the postgraduate students of Ionian University Informatics, Greece. On the basis of reviewed literature, they considered Gender, Age, Marital Status, Number of children, Occupation, Job associated with computers, Bachelor, Another master, Computer literacy, Bachelor in informatics. Mashael A. Al-Barrak and Mona S. AlRazgan collected a dataset of student's from the Information Technology department at King Saud University, Saudi Arabia for their analysis. They further used the different attributes for the prediction like student ID, student name, student grades in three different quizzes, midterm1, midterm2, project, tutorial, final exam, and total points obtained in the Data structure course of the computer science department [8]. Edin Osmanbegović and Mirza Suljic collected data from surveys in the midst of first-year students and the data taken during the enrollment at the University of Tuzla. They further used the different attributes for the prediction like Gender, Family, Distance, High School, GPA, Entrance exam, Scholarships, Time, Materials, the Internet, Grade importance, Earnings. Raheela Asif and Mahmood K. Pathan in their study used data from four academic batches of Computer Science & Information Technology (CS & IT) department at NED University, Pakistan. They used HSSC marks , Maths marks in HSSC , and marks in programming courses like Logic design, OOP, DBMS and Data Structures .

To conclude , after the review of different research papers it was found that in most of the cases the factors which affect the student performance are gender, high school grade, student's parental education, financial background, living location, medium of teaching, student's family status, students' previous semester marks, class test grade, seminar performance, assignment performance, general proficiency, attendance in class and lab work, interest in particular course, admission type and previous schools marks .

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 Functional Requirements

Academic institutions need to create an admin account and login in order to access the functionalities on the portal. Admin will import their student's data for processing , performing data analytics and prediction. Data can be visualised in the form of bar graphs, box plots etc. Admin or user can view the desired subset of the data using filters. Users can perform data analytics on past academic data to make useful insights. The visual output will neatly show the predictions on the data. Users can view the calculations and performance metrics of different models.

3.2 Non Functional Requirements

3.2.1 Performance Requirements

The web server should be able to handle multiple users in parallel, also the response time should not be too long and the server should provide periodic notification regarding the progress. The system expects that the user will have a standard Internet connection. Since the application depends on third parties, also the system assumes that third party services will also be consistent.

3.2.2 Safety Requirements

Backup is maintained for safety purposes. The damage to the data and the server is provided by regular surveillance of the system. This is to save the system from any malicious activities that can be carried out. The web server and database are expected to have enough capabilities to prevent data breaches and to provide periodic backups. User profiles shall not be disclosed to any third party, and minimal pertaining information of the user will be kept.

3.2.3 Security Requirements

For security reasons not every user is allowed to access all the data. Each user is provided with a login user id and password to protect the data. By providing login and password, unauthorized access to data cannot be made. The application should communicate with a remote server using a secure socket protocol. This ensures the confidentiality, integrity, and non-repudiability of the payload. Furthermore, all data on the database shall be encrypted and third parties would not be allowed to persist or distribute data.

3.2.4 User Documentation

The application should include elaborate and intuitive controls that will enable the user to use the application without expert knowledge of the domain. However, instructions should be included in the application for completeness's sake. A complete document will be provided to an institute about the proper functioning of the system. It will include the conditions and constraints. The technologies such as the frontend and the backend will also be defined.

CHAPTER 4

METHODOLOGY AND IMPLEMENTATION

4.1 Process Diagram

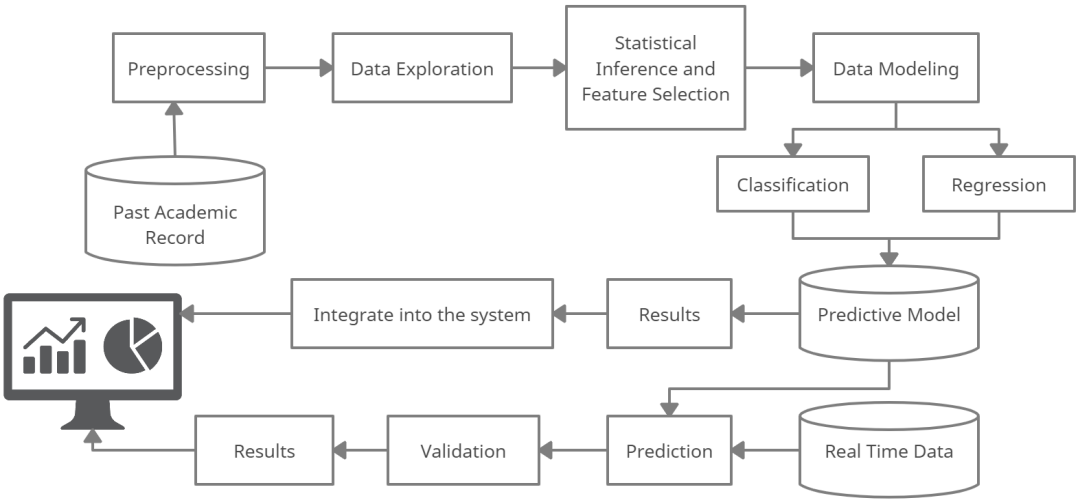


Figure 1. Process Diagram

4.2 Data collection

The data used for our project was provided by FAST NUCES . The data contained academic records of undergraduate level (Bachelors) students for the past 19 years from Fall 2001 to Summer 2019 . The data was provided for all the FIVE campuses of NUCES i-e Karachi , Lahore , Peshawar , Faisalabad and Islamabad. The dataset provided was given in four separate excel sheets Student Data , Semester Data , Course Data 1 , Course Data 2 .

- **Student Data :**

  This data set contained all the relevant detail about a particular student i-e gender , batch , campus , program code , CGPA , first semester , last semester , city , SSC Board , SSC obtained , SSC Total , HSSC Board , HSSC obtained , HSSC Total, O Level Board , O Level Obtained , O Level Total , A Level Board , A Level Obtained , A Level Total , warnings , credits attempted , credits completed. This all was given against a unique student id.

- **Semester Data:**

  This data set contained the academic details of students for each semester throughout the graduation cycle. The attributes included semester , sgpa , cgpa , core course count , elective course count . Information about each semester of a particular student was

given row wise i-e for a single student there will be multiple rows each for a particular semester.

- **Course Data 1 & Course Data 2 :**
  Both these datasets had the same columns: semester , student id , code , title , credit hours , course type , relation id , grade , grade point . The total number of students were split into two halves , one half was in Course Data 1 and the other in Course Data 2 . Each row showed data about a particular course of a particular student i-e data of each student was given in several rows to cover all his/her courses.

4.3 Data Preparation

Whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. Therefore the data has to go through pre-processing in which it is transformed (as needed) into a form which can be easily used for the analysis. Since our dataset was provided in four different files , we pre processed each file accordingly .

- **Student Data**

  Columns such as warnings , credits attempted , credits completed , SSC Total , HSSC total were dropped. A single column was made for Secondary Education from the SSC Board and O Level Board . Also SSC Obtained and O Level Obtained were combined to form a column Secondary Grade. The columns SSC Board and O Level Board were combined to form a column School. Similarly , a single column was made for Higher Secondary Education from the HSSC Board and A Level Board . Also HSSC Obtained and A Level Obtained were combined to form a column Secondary Grade. The columns HSSC Board and A Level Board were combined to form a column College.

- **Semester Data**

  For sorting the data the semester attribute was splitted into year and session. The data was transformed into a new dataframe in which each row had a unique student id against which there were columns for sgpa and cgpa from the first to the last semester. Elective Course Count and Core Course Count were dropped as they were not of any use for our analysis.

Finally above two data sets were joined on unique student ids to form a single data set.

- **Course Data 1 & Course Data 2**

  From course code which was given like SS123 the course domain i-e SS was extracted. From the relation id attribute only core courses were retained and the elective courses were dropped , as they were not a part of our FYP scope. Columns that weren't useful were dropped and only columns student id , title , domain and grade point were kept for further work . The courses were then splitted domain wise i-e CS , EE , SS , MG , CV , MT , EL , CL , VL, FYP. The dataset was transformed in a way that all courses were placed column wise and separate sheets were maintained for each domain to find out relation between different courses of the same domain.

Point to be noted is that we chose only those courses which were prerequisites of some other courses , because for finding relationships between courses we only needed the courses in chain.

Once that has been pre processed , it has to be cleaned against inconsistencies and missing values. In our data set we found a lot of missing data and a few inconsistencies.

Missing Values in Dataset

```
##       STUDENT_ID       SEM_1_SGPA       SEM_1_CGPA       SEM_2_SGPA
##                0             5783             5786            10812
##       SEM_2_CGPA       SEM_3_SGPA       SEM_3_CGPA       SEM_4_SGPA
##             7628            12977            12081            15550
##       SEM_4_CGPA       SEM_5_SGPA       SEM_5_CGPA       SEM_6_SGPA
##            13571            16788            15994            18509
##       SEM_6_CGPA       SEM_7_SGPA       SEM_7_CGPA       SEM_8_SGPA
##            16859            19377            18690            20947
##       SEM_8_CGPA       SEM_9_SGPA       SEM_9_CGPA      SEM_10_SGPA
##            19469            28122            27433            32003
##      SEM_10_CGPA      SEM_11_SGPA      SEM_11_CGPA      SEM_12_SGPA
##            31549            34667            34389            36599
##      SEM_12_CGPA      SEM_13_SGPA      SEM_13_CGPA      SEM_14_SGPA
##            36398            37748            37598            38473
##      SEM_14_CGPA      SEM_15_SGPA      SEM_15_CGPA        TOTAL_SEM
##            38372            38807            38751                0
##           GENDER            BATCH           CAMPUS        PROG_CODE
##                0                0                0                0
##             CGPA        FIRST_SEM         LAST_SEM           STATUS
##             5572                0                0                0
##             CITY        SECONDARY           SCHOOL        SEC_GRADE
##                0             1195             5587             1195
## HIGHER_SECONDARY          COLLEGE    HIG_SEC_GRADE
##             4481            10506             4481
```

Figure 2. Missing Values in Dataset

The duplicate student ids within rows were removed from the dataset . To cater inconsistencies in school name and college name , upper casing was done and extra spaces were removed. To cater null values in categorical variables such as school name , college name , secondary , higher secondary row removal was done. For numerical attributes such as the sgpa , secondary grade , higher secondary grade mean imputation was done . For cgpa , to fill null values , the proper cgpa calculation was done using spga. Columns for sgpa and cgpa of semester above 8 were dropped , since most of the values in the column were null. Mean imputation was also done to fill missing values of grade points of courses.

4.4 Statistical Analysis

Statistical analysis is the collection and interpretation of data in order to uncover patterns and trends. This is the most important part of our project .Our data had both numerical and categorical variables. Therefore , while finding correlations we applied appropriate techniques for numerical vs numerical and numerical vs categorical respectively. We applied the following statistical techniques to perform our analysis:

1. **Analysis of Variance (ANOVA)**

   Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among means.

For numerical feature vs categorical feature we have used One Way ANOVA. In this we take a null hypothesis and an alternate hypothesis. By applying ANOVA we get a p-value that is used to accept or reject the null hypothesis. If the p value is less than significance value we reject the null hypothesis.

2. **Pearson correlation**

For numerical vs numerical we have used Pearson Correlation Method which not only provides us the correlation coefficient but also the significance value , which tells us how reliable the result is. Value of correlation coefficient nearer to 1 shows highly positive correlation , nearer to -1 shows highly negative correlation whereas 0 shows no correlation.

3. **Recursive Feature Elimination**

RFE is a wrapper-type feature selection algorithm. This is achieved by fitting the given machine learning algorithm used in the core of the model, ranking features by importance and discarding the least important features

4. **Feature Importance Ranking Measure**

Feature Importance Ranking Measure (FIRM) by retrospective analysis of arbitrary learning machines allows to achieve both excellent predictive performance and superior interpretation. In contrast to standard raw feature weighting, FIRM takes the underlying correlation structure of the features into account. Thereby, it is able to discover the most relevant features .

5. **Stepwise Backward and Forward Feature Selection**

Stepwise methods have the same ideas as best subset selection but they look at a more restrictive set of models. Between backward and forward stepwise selection, there's just one fundamental difference, which is whether you're starting with a model: with no predictors (forward) with all the predictors.

4.5 Model Building

Since our project focused on exploring different machine learning algorithms for the problem of predictive analytics on academic record , therefore we treated the problem both as regression and classification . Following this we implemented different regression and classification models. The data was split into train , validation and test sets respectively . The train set was used to train the model , validation was used to check model performance and the test set (the data of currently enrolled students) was utilized for prediction .

4.5.1 Regression Models

The following models were used for the prediction of CGPA :

1. **Linear Regression**

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.

2. **Polynomial Regression**

   In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x.

4.5.2 Classification Models

For the purpose of classification , CGPA was divided into different intervals / regions. The following models were used to classify the CGPA interval :

1. **Decision Tree**

   Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

2. **Gaussian Naive Bayes**

   Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data.It is a simple classification technique, but has high functionality.

3. **K Nearest Neighbors**

   The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

4. **Logistic Regression**

   Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable. In statistics, the logistic model is used to model the probability of a certain class or event existing.

5. **Random Forest**

   A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

6. **Support Vector Machine**

In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.

7. **Ensemble Classification**

   An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples

4.6 Experimental Setup

4.6.1 Programming Language

R language was used to perform all the work related to data analytics. R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

Python was used for the work related to model building and evaluation. Python is a high level general purpose language, used by data scientists and developers, which makes it easy to collaborate across your organization through its simple syntax. People choose to use Python so that they can communicate with other people. The other reason is rooted in academic research and statistical models

4.6.2 Software Tools

For the data cleaning , transformation and EDA , R studio was used . RStudio is an integrated development environment for R, a programming language for statistical computing and graphics.

For model building and predictions , Jupyter Notebook was used. The Jupyter Notebook is an incredibly powerful tool for interactively developing and presenting data science projects.

For the development of the dashboard MERN (MongoDB, Express.js, React.js, Node.js) stack was used. It is a free and open source javascript software stack for building dynamic websites and web applications.

CHAPTER 5

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis is one of the crucial steps in data analytics. Before we jump to learning and modeling the data , EDA was to be performed. In our case , the EDA was first performed for all the data as a whole (all FIVE campuses together) and then separately for each campus .

5.1 Visualization of the whole data

As it was already mentioned that the data provided was for FIVE different campuses , the first bar chart shows the number of students in each campus .

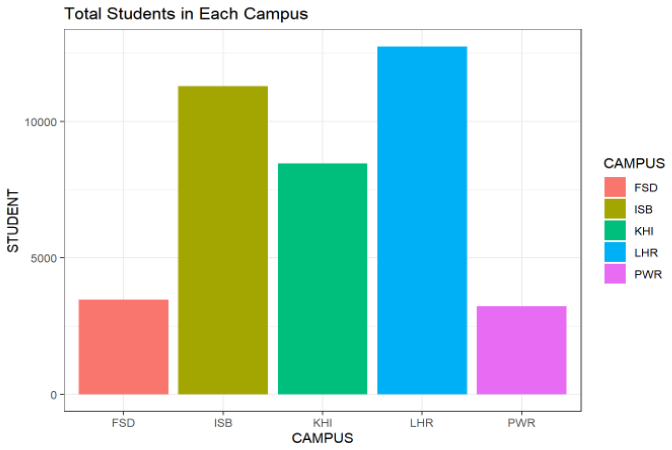


Figure 3. Total Students in Each Campus

This is quite clear from Figure 3 that for the past 19 years , out of all the five campuses the greatest number of students were enrolled in Lahore Campus .

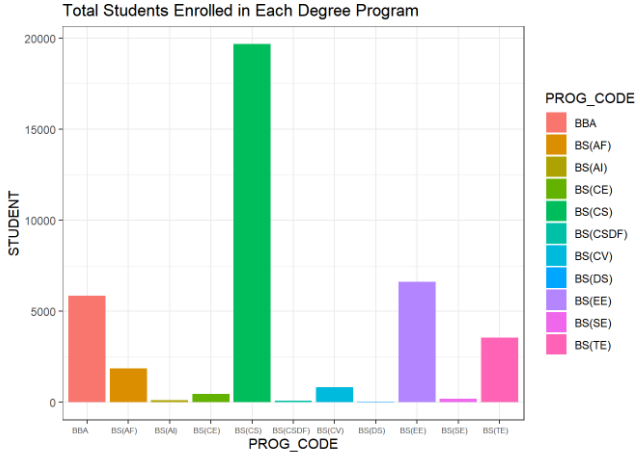


Figure 4. Total Students enrolled in each Degree Program

From Figure 4 , for the past 19 years , 11 different degree programs have been offered . Being the best Computer Science university in Pakistan , the majority of the students at FAST are enrolled in BS(CS) - Bachelors in Computer Science.

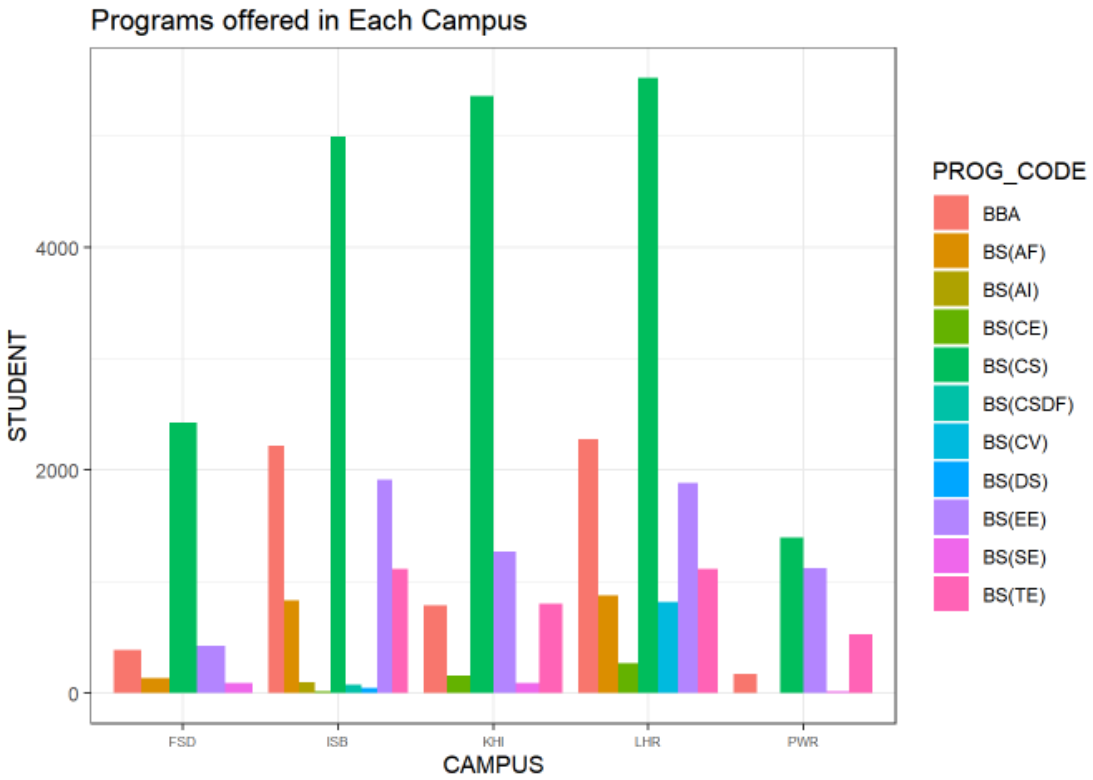


Figure 5. Programs offered in each Campus

The dataset was provided with an attribute of student status which had 7 different categories.

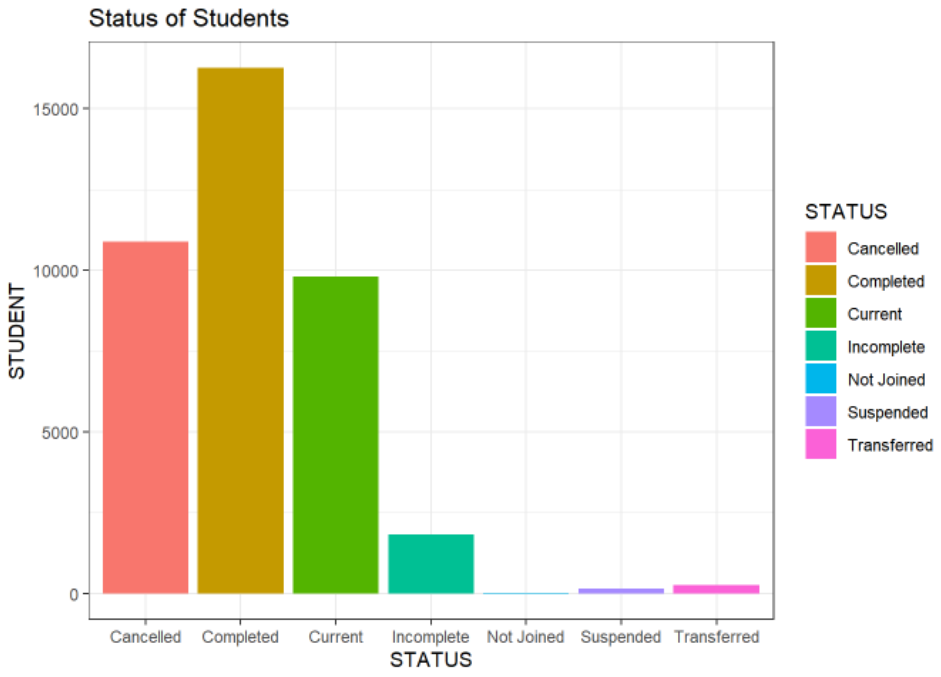


Figure 6. Degree Status of Students

The count of students with incomplete , not joined , suspended and transferred is negligible as compared to the others therefore these weren't used for the purpose of analysis. Also the students with cancelled status , didn't complete their graduation and not enough data was available about them they were also excluded from the analytics .

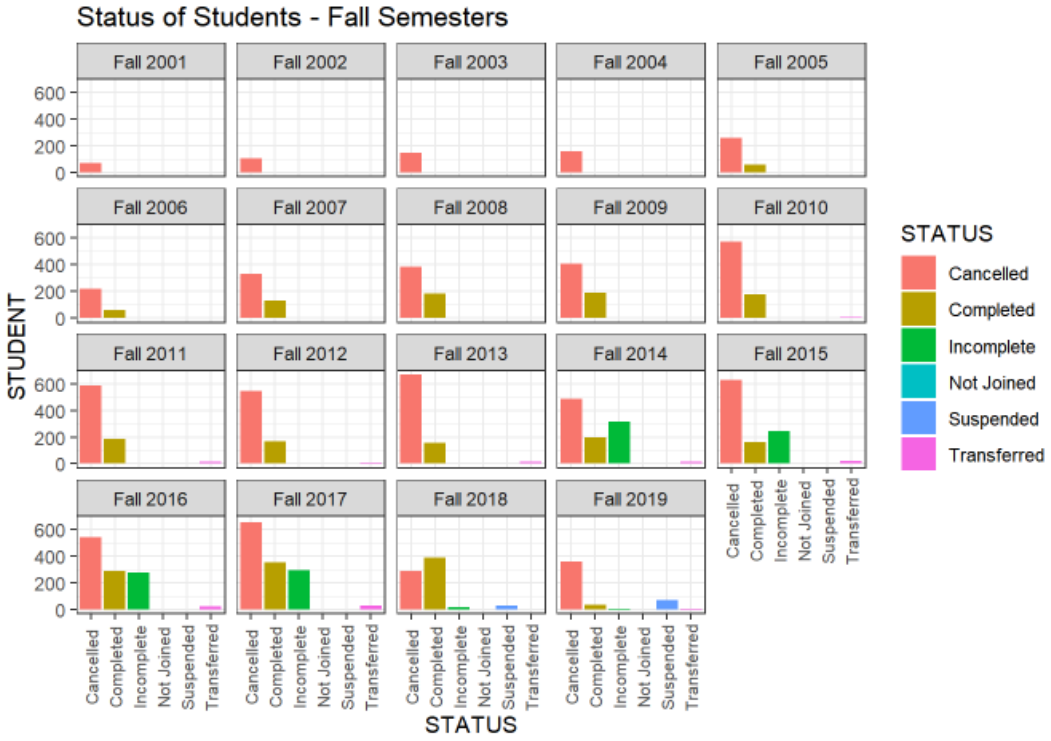


Figure 7. Degree Status of Students - Fall Semester

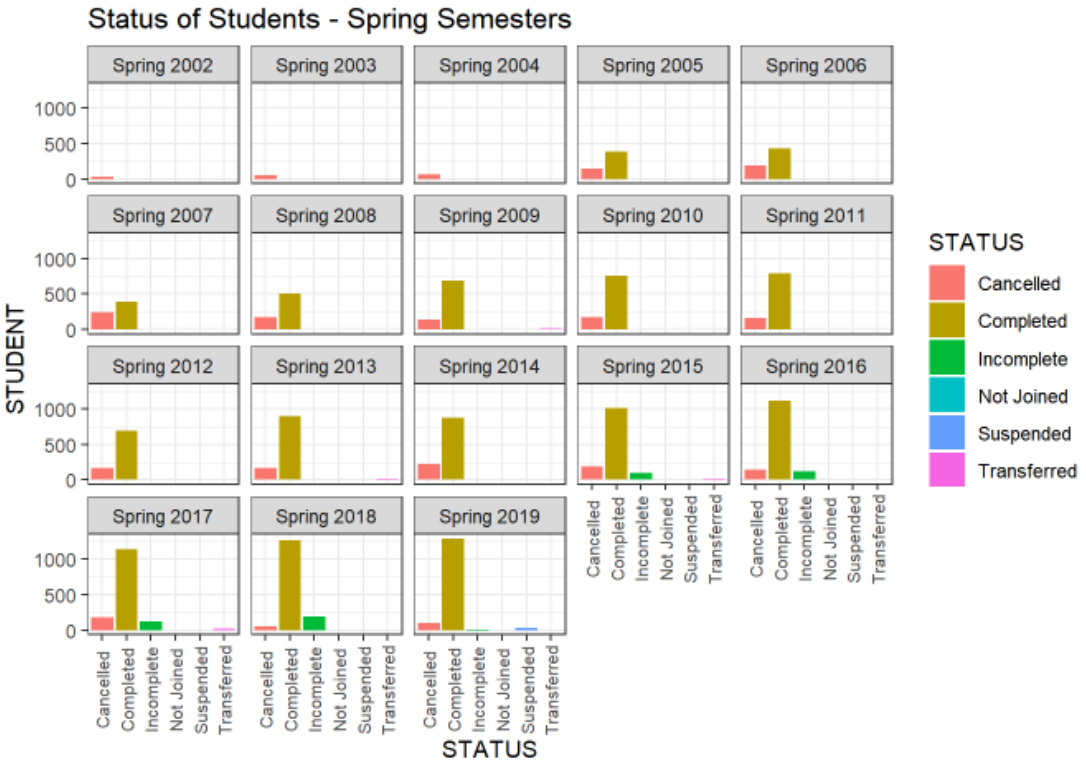


Figure 8. Degree Status of Students - Spring Semester

Figure 7 and 8 are the visualization of students' status for different semesters of Fall and Spring for the past 19 years.

Figures 9 and 10 are clear evidence of an imbalance of male and female students. Around 81.22% students enrolled in five campuses of FAST NUCES around the country are found to be males.

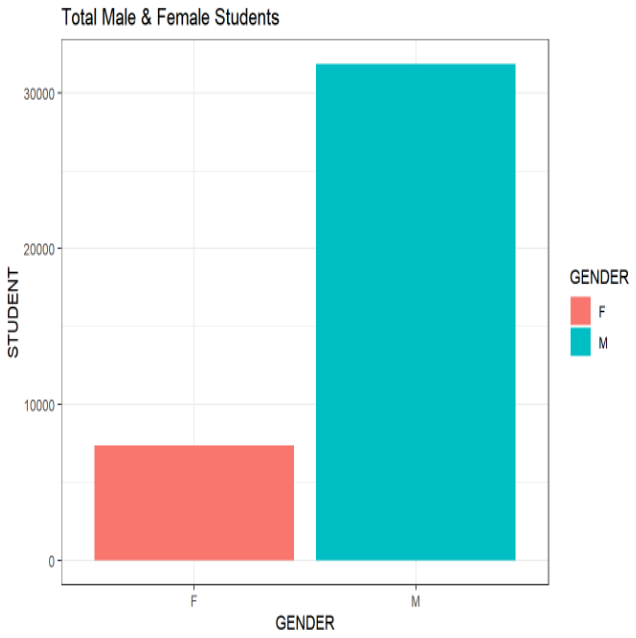


Figure 9. Total Male & Female Students

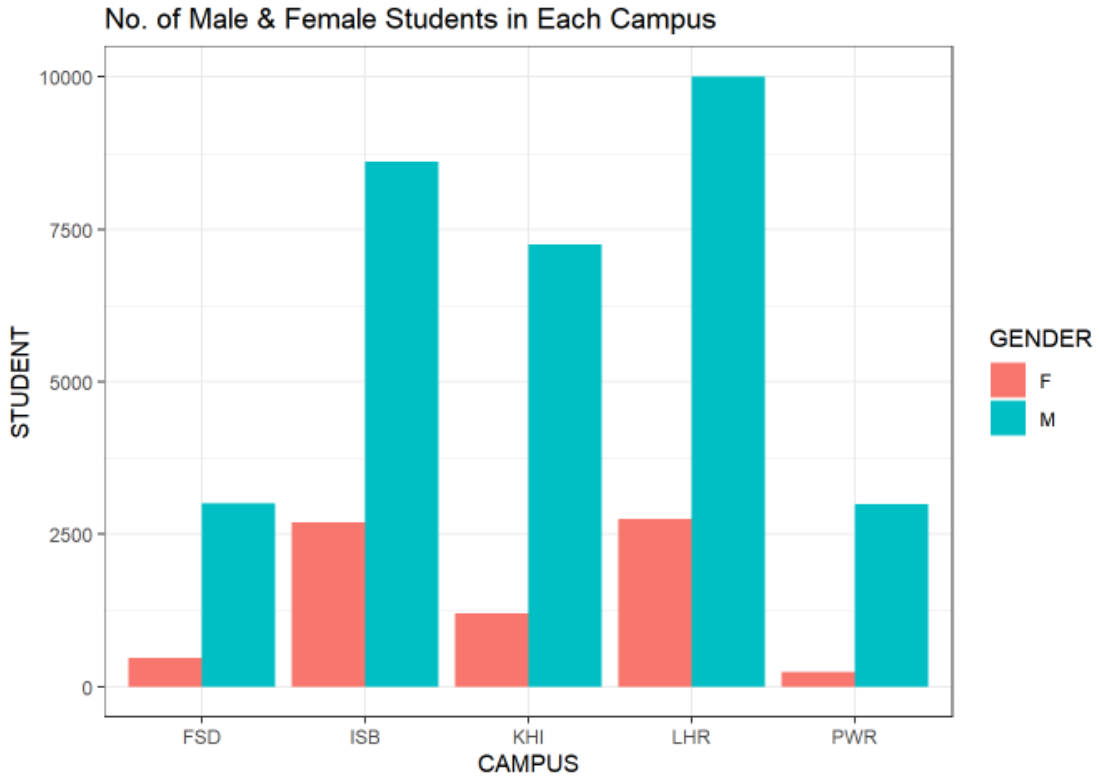


Figure 10 . Male & Female students in each Campus

The students enrolled in different campuses of FAST NUCES are from 151 different cities around the country . Some of the cities have a great majority of students but there are some who have quite negligible student count .

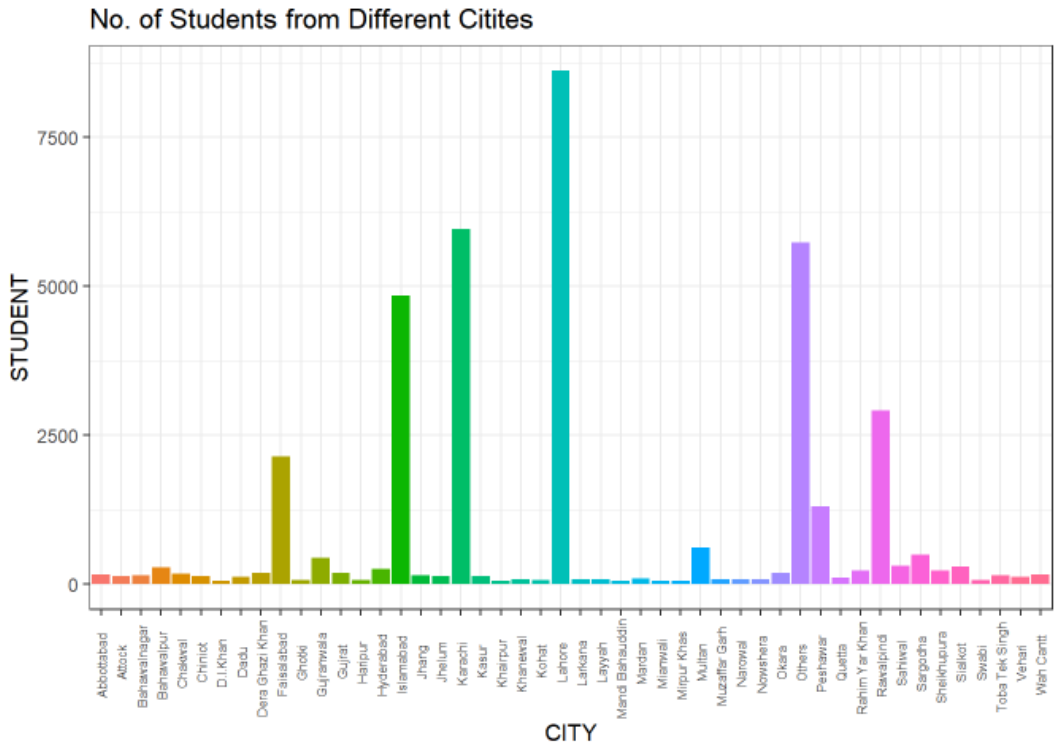


Figure 11. Number of students from different cities

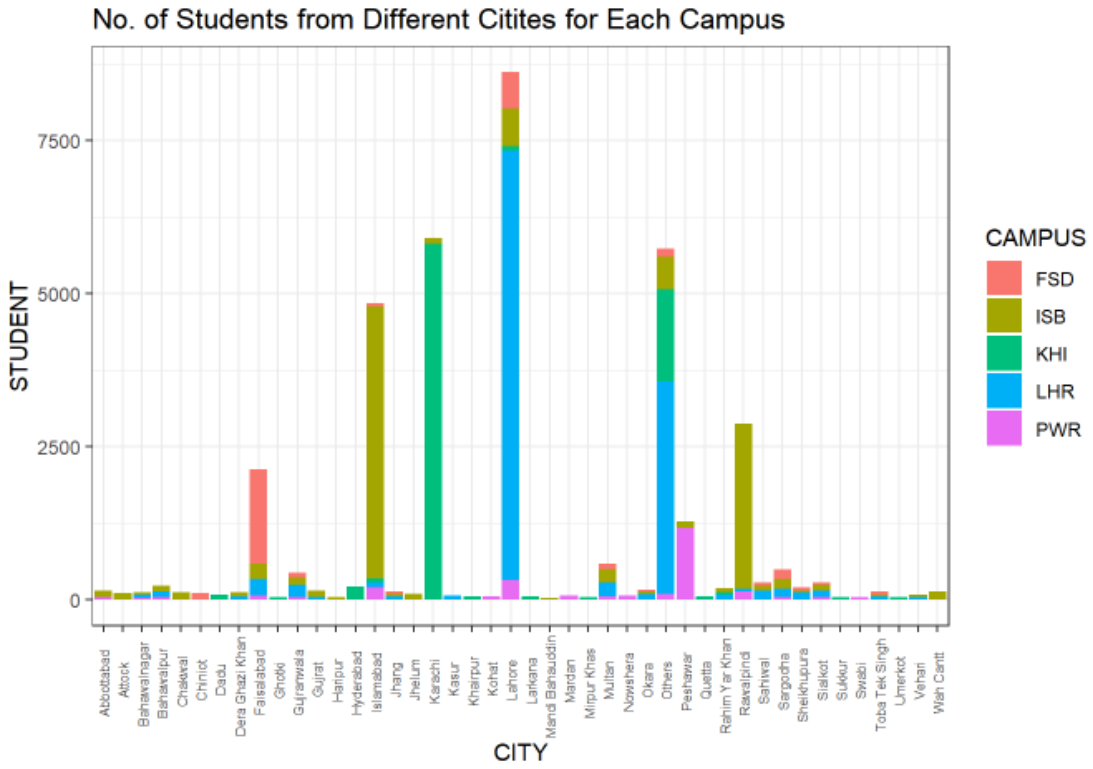


Figure 12. Number of students from different cities for each campus

For better visualizations after a threshold of min 50 students , top 48 cities are displayed in both Figure 11 and 12.

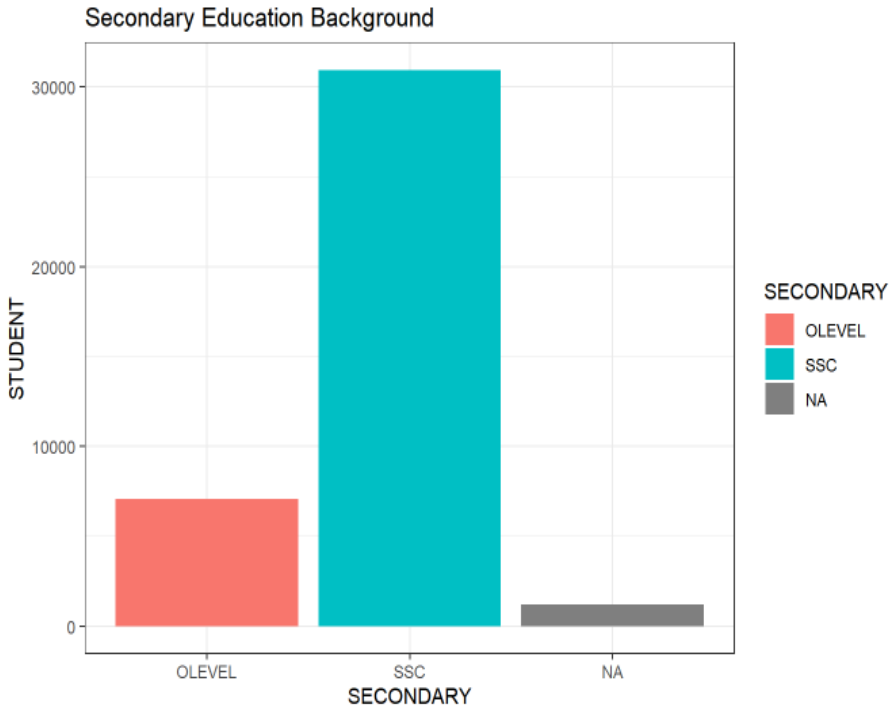


Figure 13. Status of students from Secondary Education Background

Majority of students , around 78.92% , enrolled in FAST NUCES across the country are from SSC educational background.

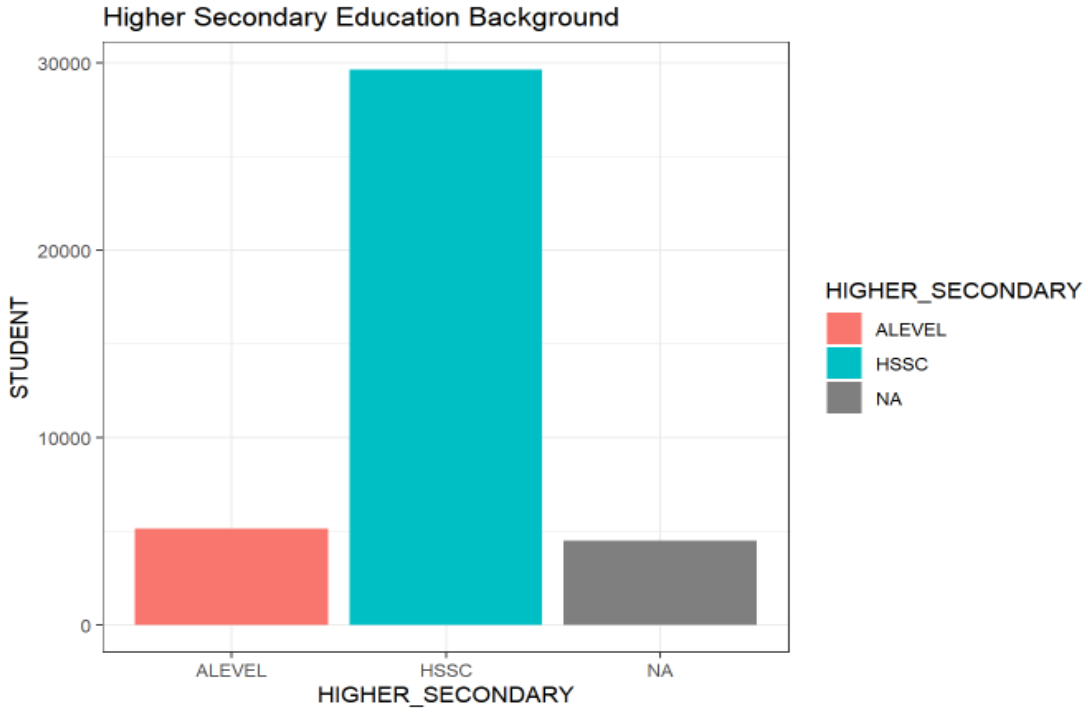


Figure 14. Status of students from Higher Secondary Education Background

Similarly , since most of the students' secondary education is from SSC background , the higher secondary education of most students , around 75.5% , is from HSSC background.

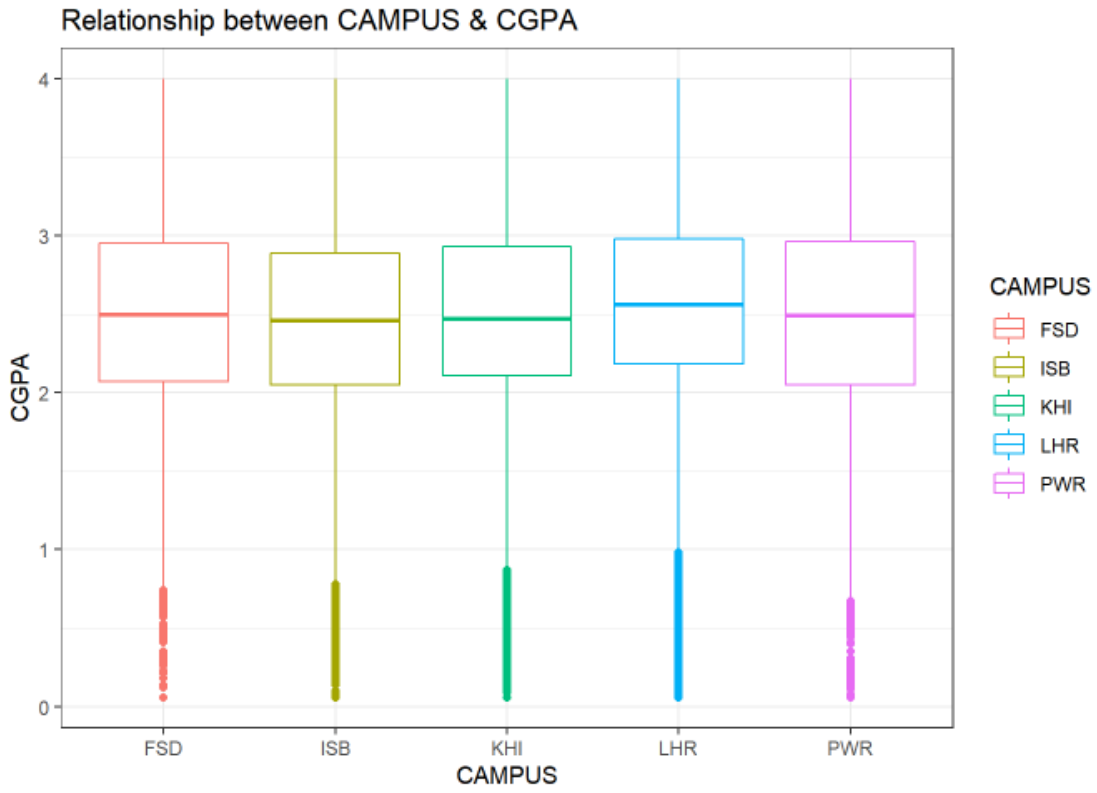


Figure 15. Boxplot of CGPA for each campus

Figure 15 visualizes the campus wise CGPAs. It can be seen that in each campus there are outliers . However , the mean of each campus is quite the same .

COURSE_DOMAIN	count
CL	18
CN	2
CS	141
CV	38
DS	2
EE	109
EL	59
ME	3
MG	143
ML	5
MS	1
MT	34
NL	3
NS	13
SE	1
SL	5
SS	66
VL	27

Figure 16. Count of courses offered in each Course Domain

The total number of courses offered in 5 different campuses in different degree programs for the past 19 years were found to be 641. These courses were from different domains . The above below table shows total courses in each domain.

5.2 Visualization of the data of Faisalabad Campus

At Faisalabad Campus , 5 degrees were offered .

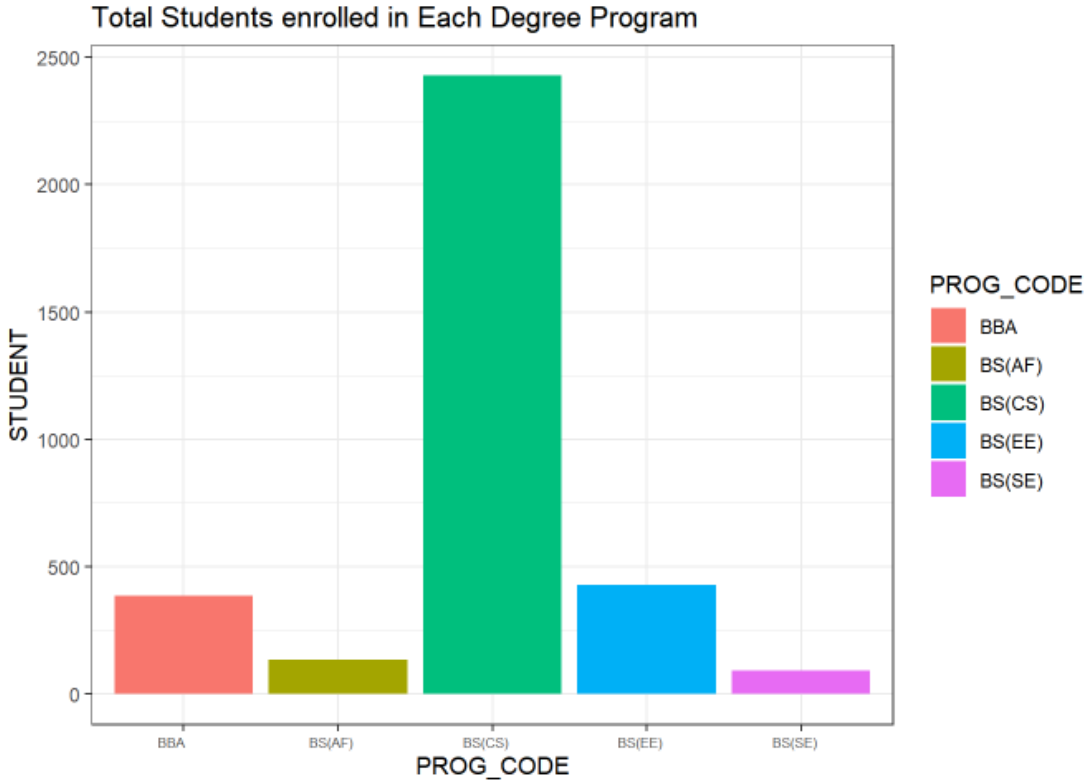


Figure 17. Total Students enrolled in each Degree Program - Faisalabad Campus

A great many students from Faisalabad are enrolled in BS(CS) whereas BS(EE) and BBA have almost the same student count over the past 19 years .

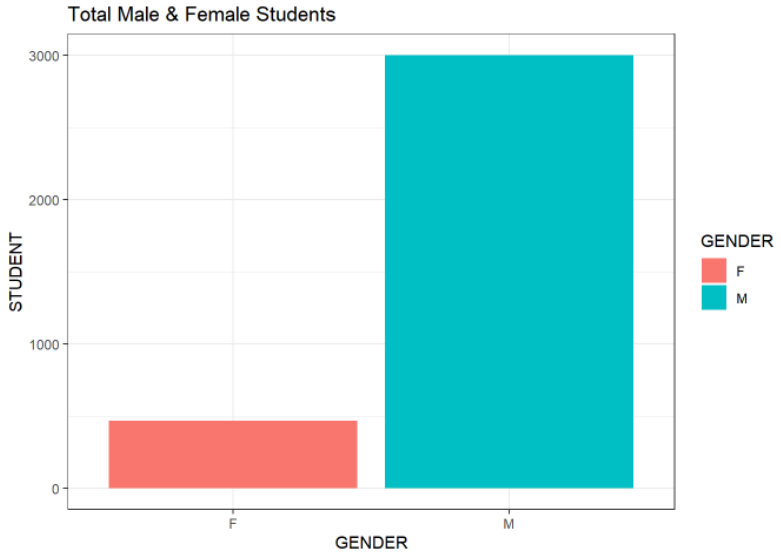


Figure 18. Total Male & Female Students - Faisalabad Campus

Figure 18 shows that from a total of 3,468 students , around 86.5% students are male .

Students enrolled in Faisalabad campus are from 49 different cities. For better visualization , after a threshold of minimum 10 students , 30 cities are displayed in Figure 19.

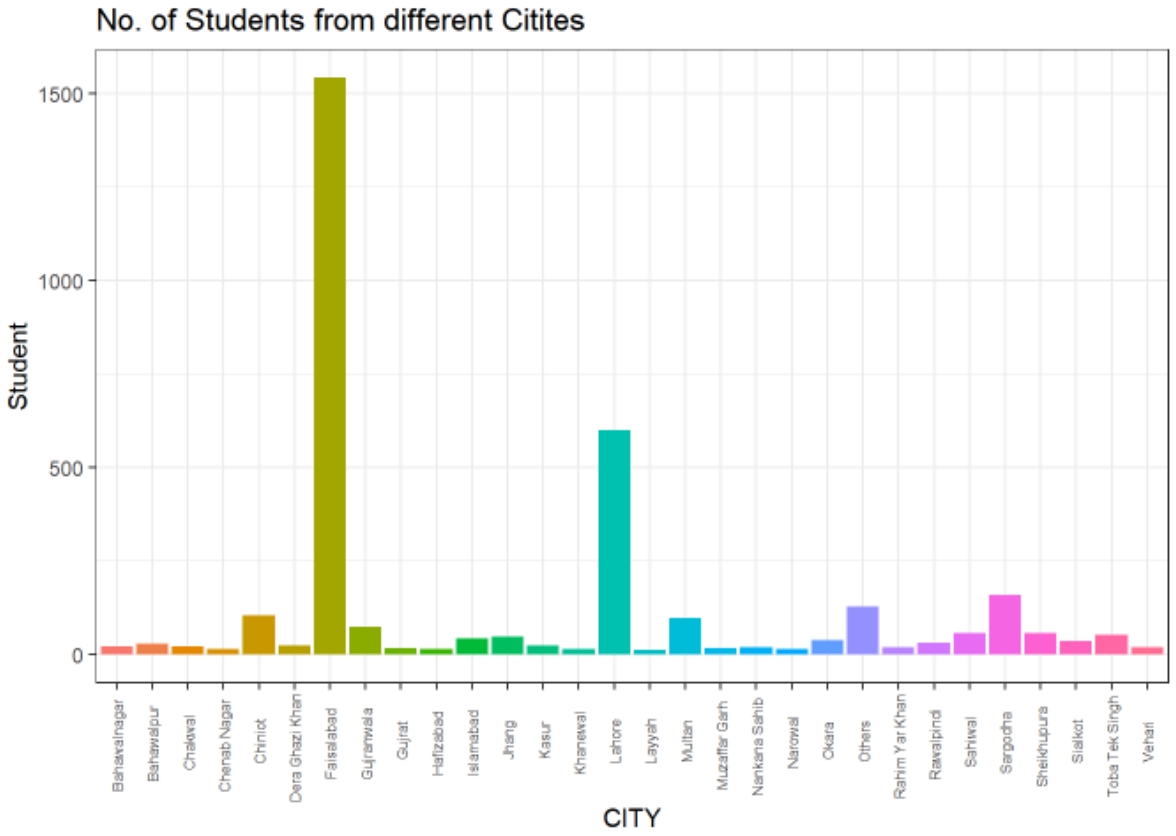


Figure 19. Number of Students from different cities - Faisalabad Campus

Majority of the students in Faisalabad Campus are from Faisalabad. However, from Lahore even there are a representative number of students.

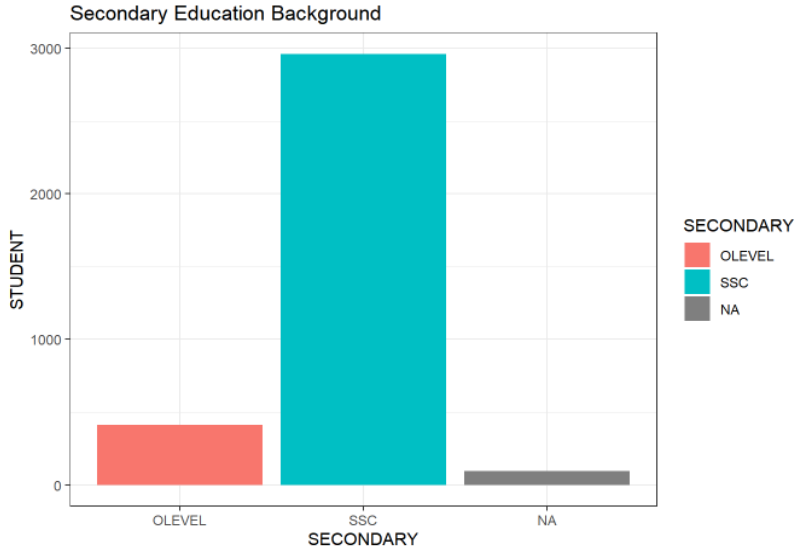


Figure 20 . Secondary Education Background - Faisalabad Campus

A great majority i-e around 85.38% students had SSC in their Secondary Education.

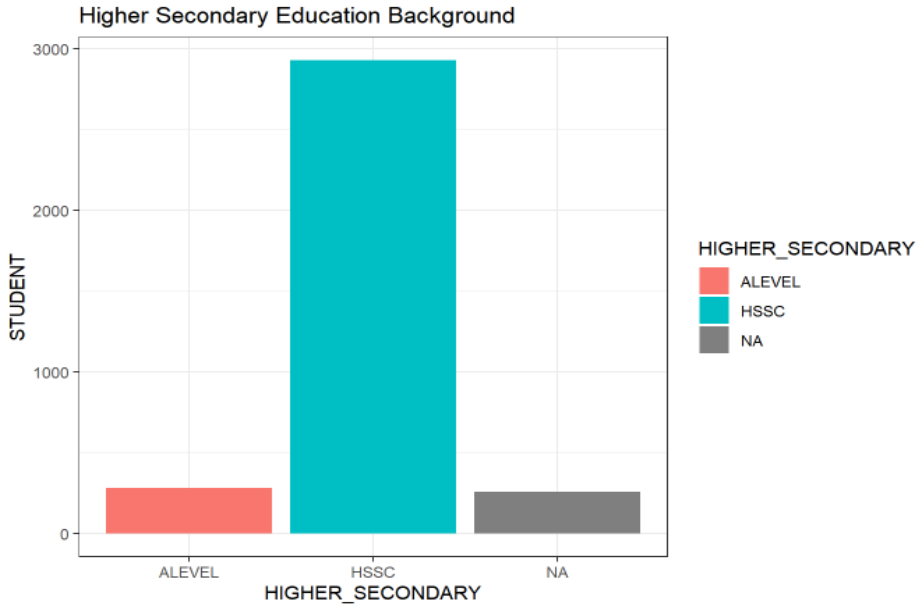


Figure 21 . Higher Secondary Education Background - Faisalabad Campus

Around 84.48% students had HSSC in their Higher Secondary Education.

5.3 Visualization of the data of Islamabad Campus

Islamabad Campus offered 9 different degree programs.

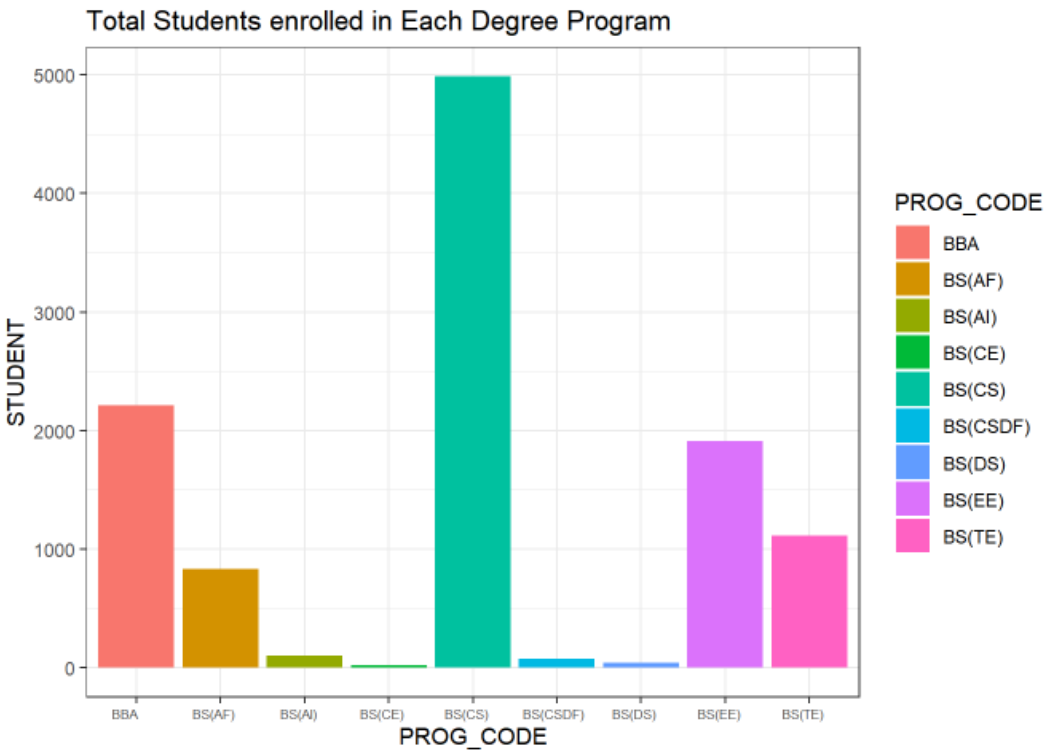


Figure 22. Total Students enrolled in each Degree Program - Islamabad Campus

The students enrolled in BS(CS) clearly exceed the other degree programs , but still a significant proportion of students have also enrolled in BBA and BS(EE).

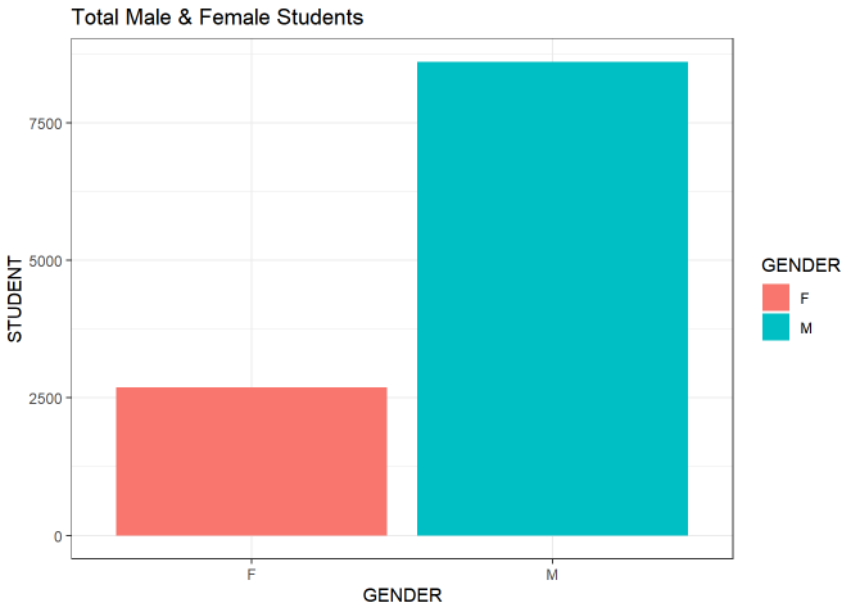


Figure 23. Total Male & Female Students - Islamabad Campus

Out of 11,305 students , 76.16% students are male at Islamabad Campus.

Students from 119 different cities around the country are enrolled at Islamabad Campus. For better visualizations , after a minimum threshold of 10 students , 57 different cities are displayed in Figure 24 .

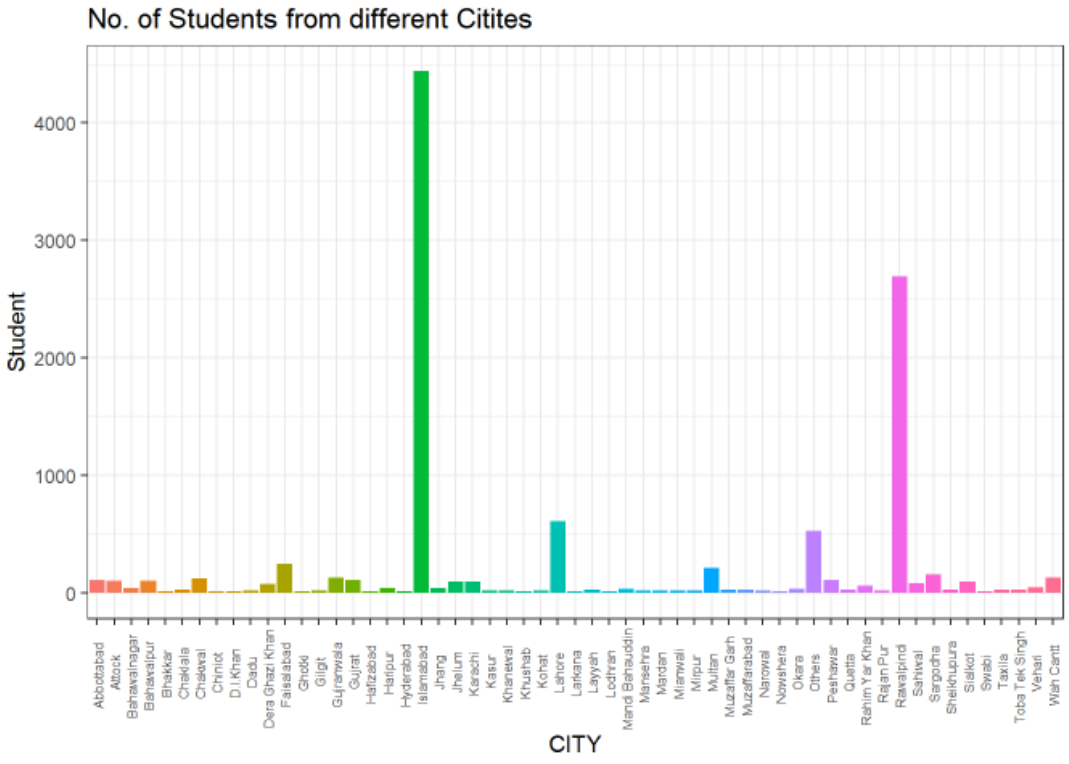


Figure 24. Number of Students from different cities - Islamabad Campus

A significant number of students enrolled in Islamabad Campus are from Islamabad and Rawalpindi.

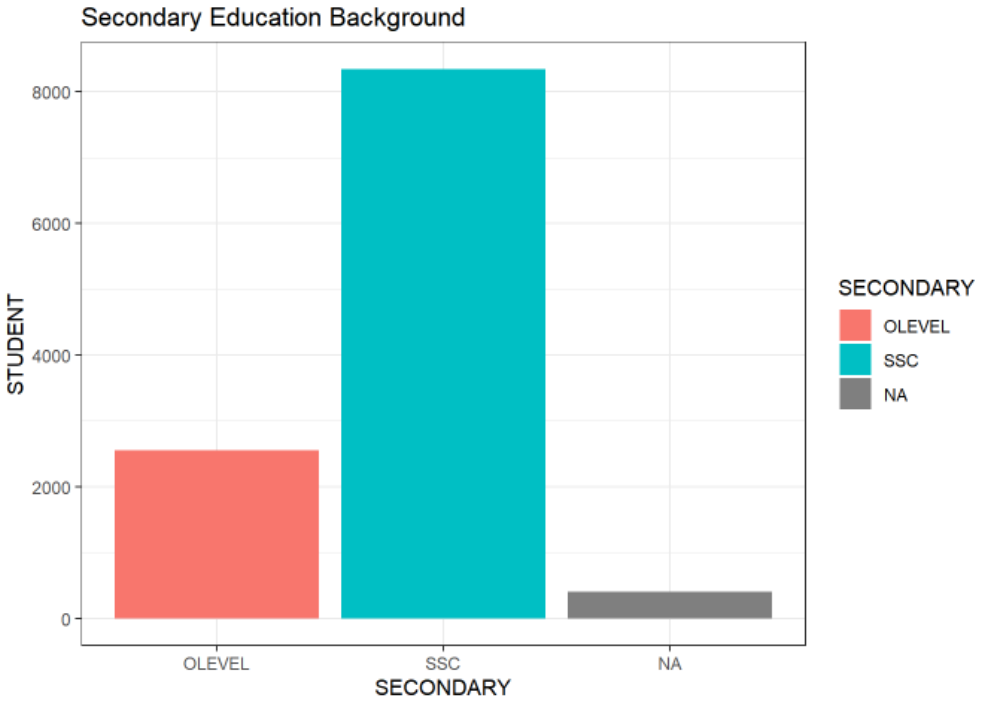


Figure 25 . Secondary Education Background - Islamabad Campus

From 11,305 students at Islamabad 73.82% students are from SSC background.

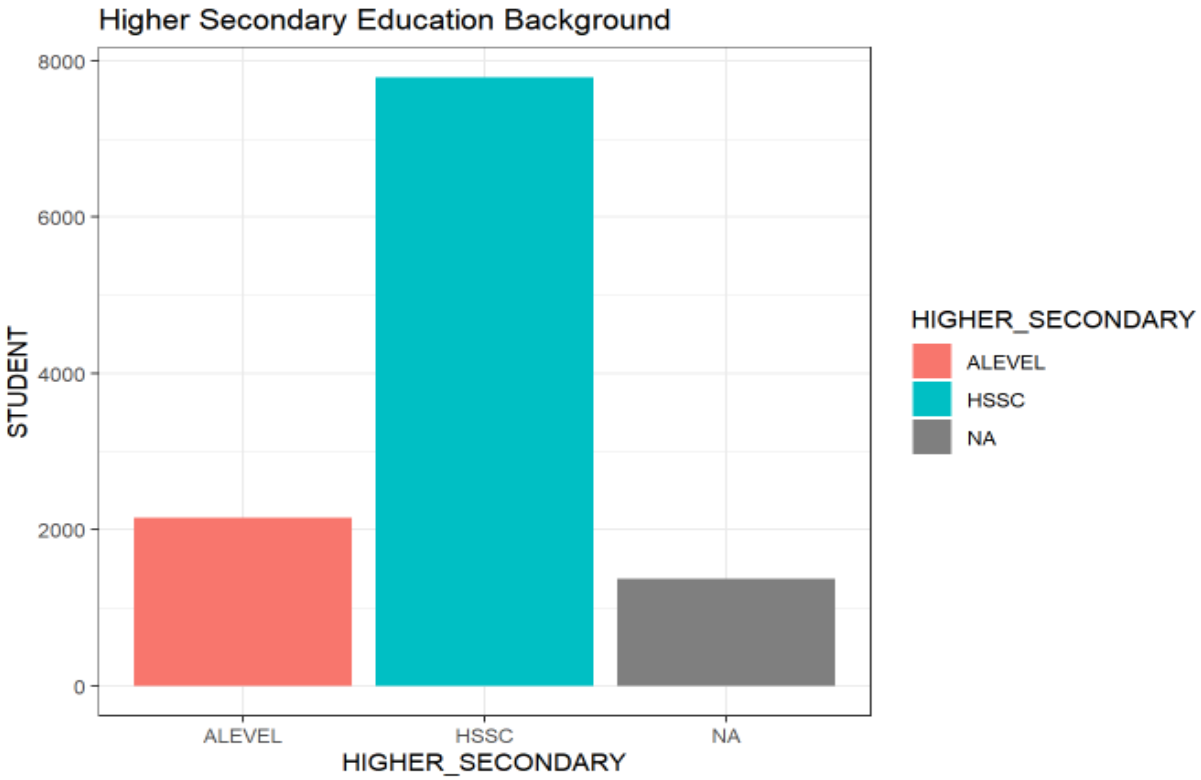


Figure 26 . Higher Secondary Education Background - Islamabad Campus

However , in Islamabad , 68.93% students are from HSSC background.

5.4 Visualization of the data of Karachi Campus

Karachi Campus offered 6 different degree programs .

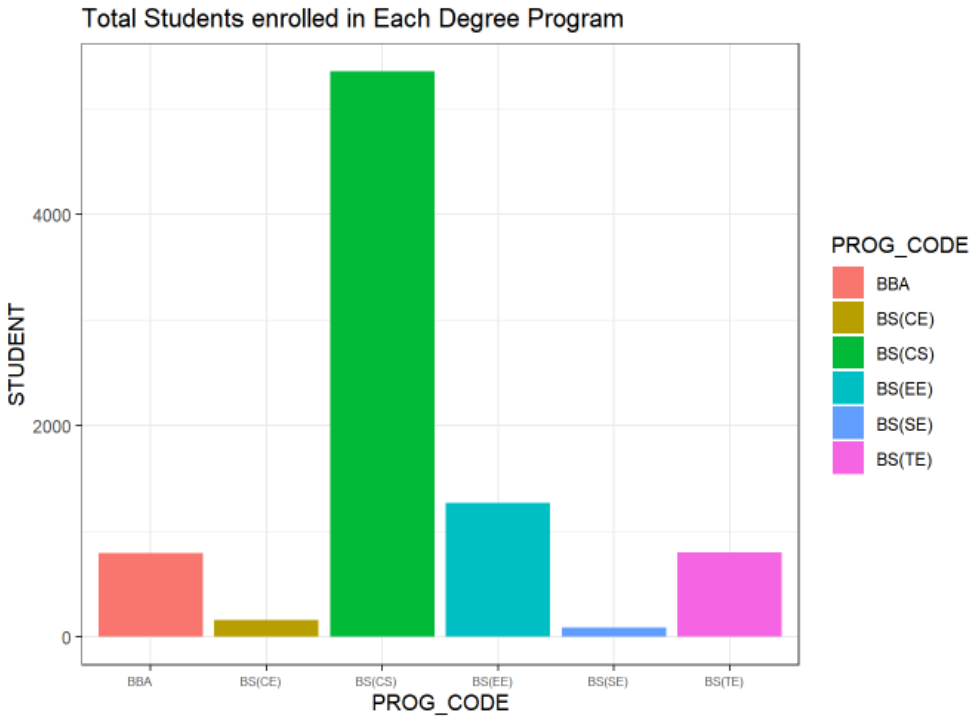


Figure 27. Total Students enrolled in each Degree Program - Karachi Campus

Out of which majority are enrolled in BS(CS) whereas students count in BBA and BS(EE) is nearly the same.

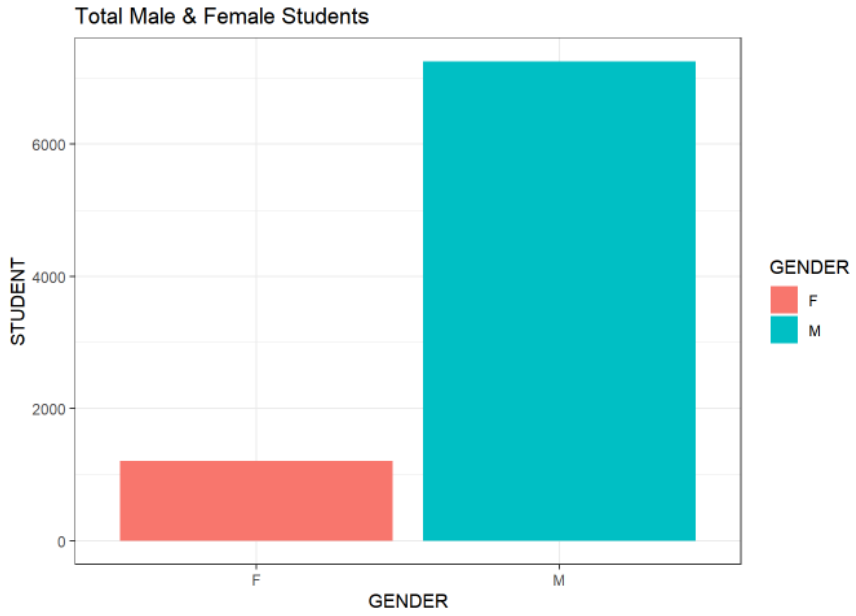


Figure 28. Total Male & Female Students - Karachi Campus

Out of 8,459 students , 85.77% students are male at Karachi Campus.

Students enrolled in Karachi Campus are from 78 different cities . However, after the threshold of minimum 10 students , 25 cities are displayed in Figure 29.

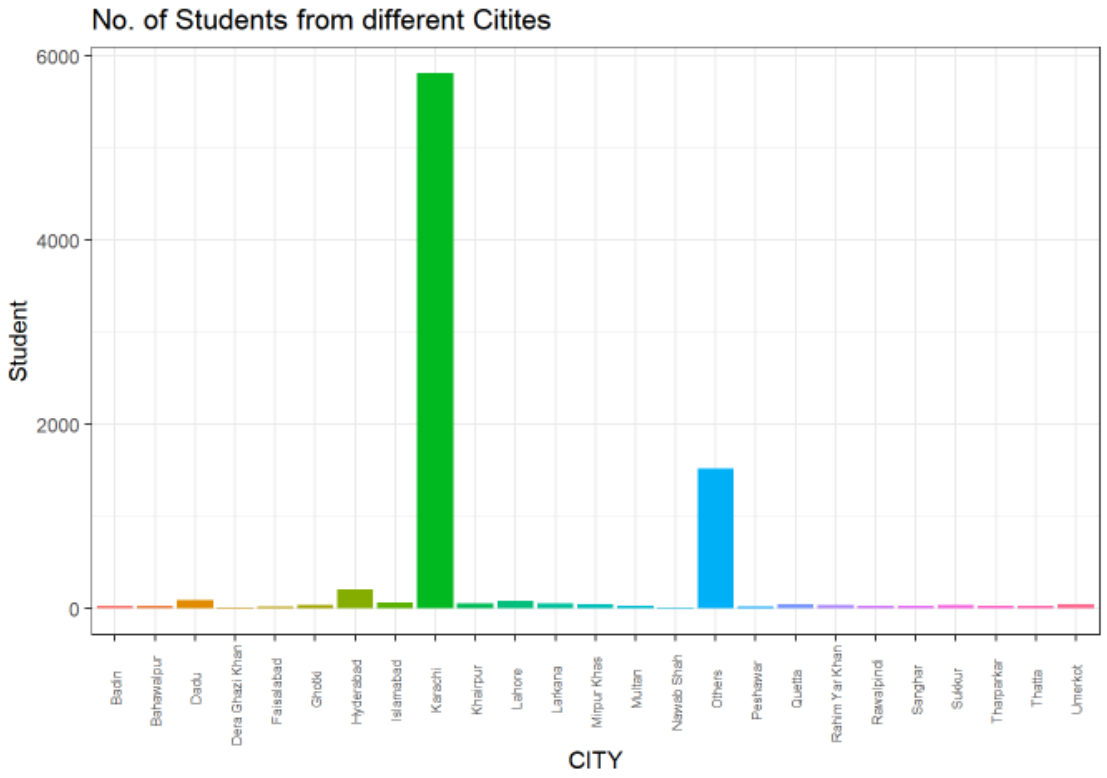


Figure 29. Number of Students from different cities - Karachi Campus

The majority of students in Karachi Campus are from Karachi .

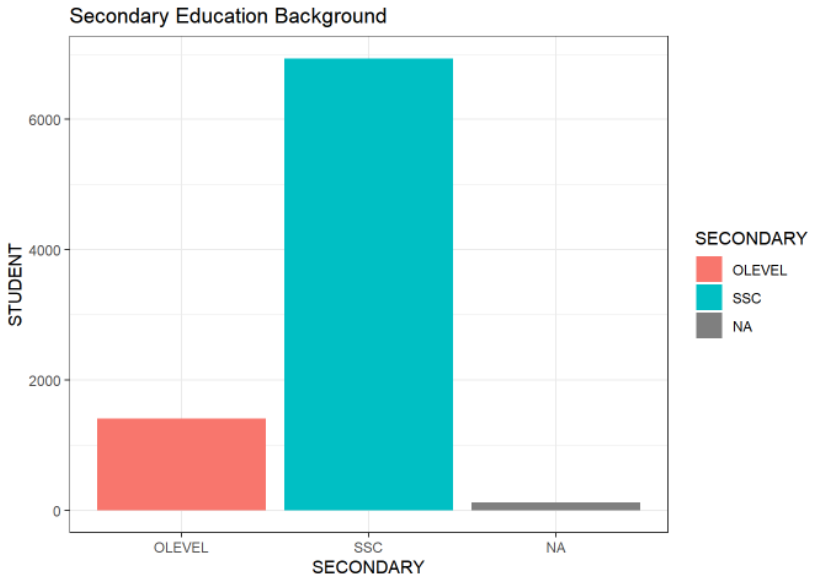


Figure 30. Secondary Education Background - Karachi Campus

At Karachi Campus , 81.99% of 8,459 students are from SSC background.

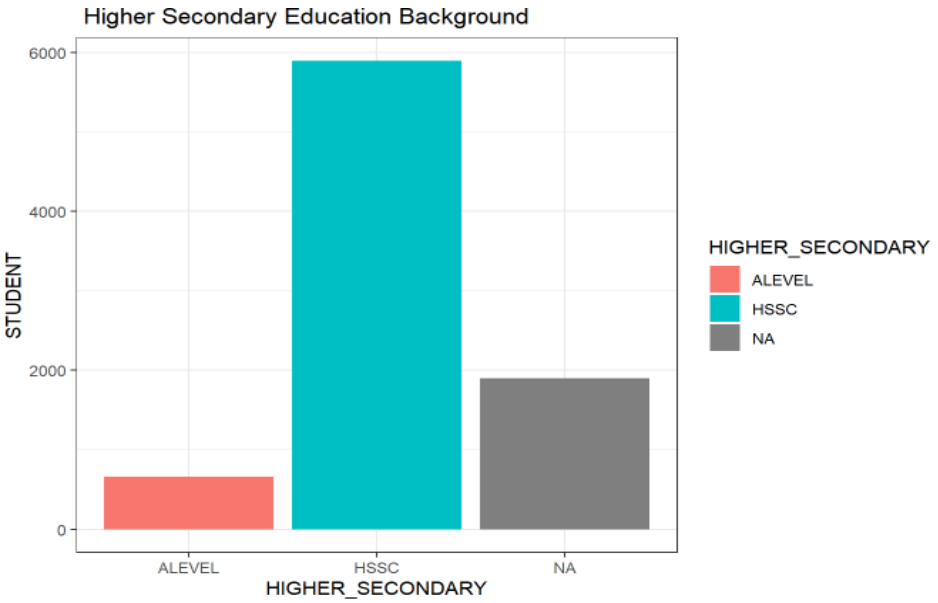


Figure 31. Higher Secondary Education Background - Karachi Campus

However , 69.74% of 8,459 students are from HSSC background.

5.5 Visualization of the data of Lahore Campus

At Lahore Campus 7 different degree programs are offered.

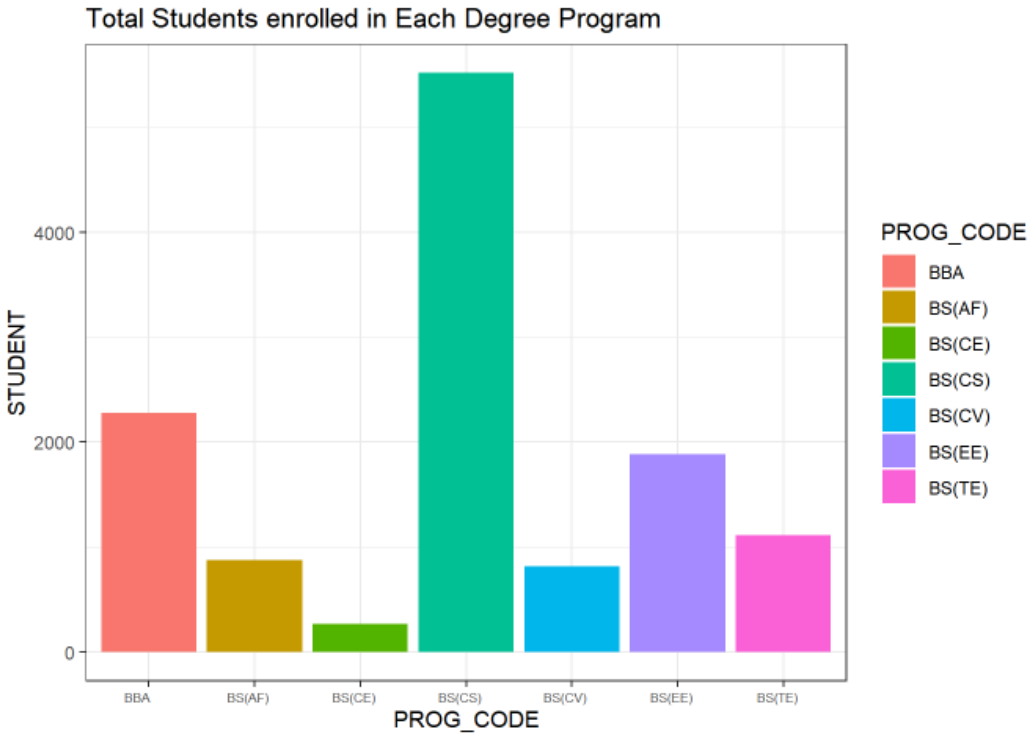


Figure 32. Total Students enrolled in each Degree Program - Lahore Campus

Compared to other students, Lahore campus has a number of students enrolled in BS(TE) and also a good number of students in BBA and BS(EE) .

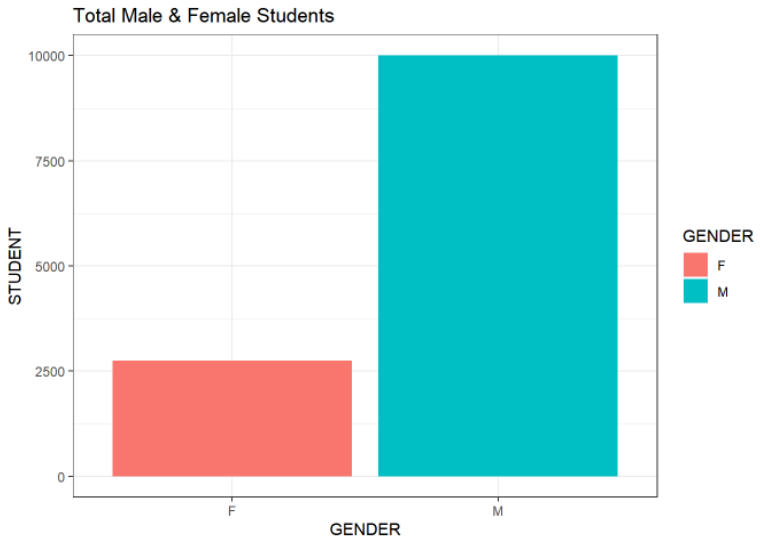


Figure 33. Total Male & Female Students - Lahore Campus

Out of 12,576 students at Lahore , male students are in a majority of 78.48% .

In Lahore Campus , students from 80 different cities are enrolled. After a minimum threshold of 10 students , 32 cities are displayed in Figure 34 .

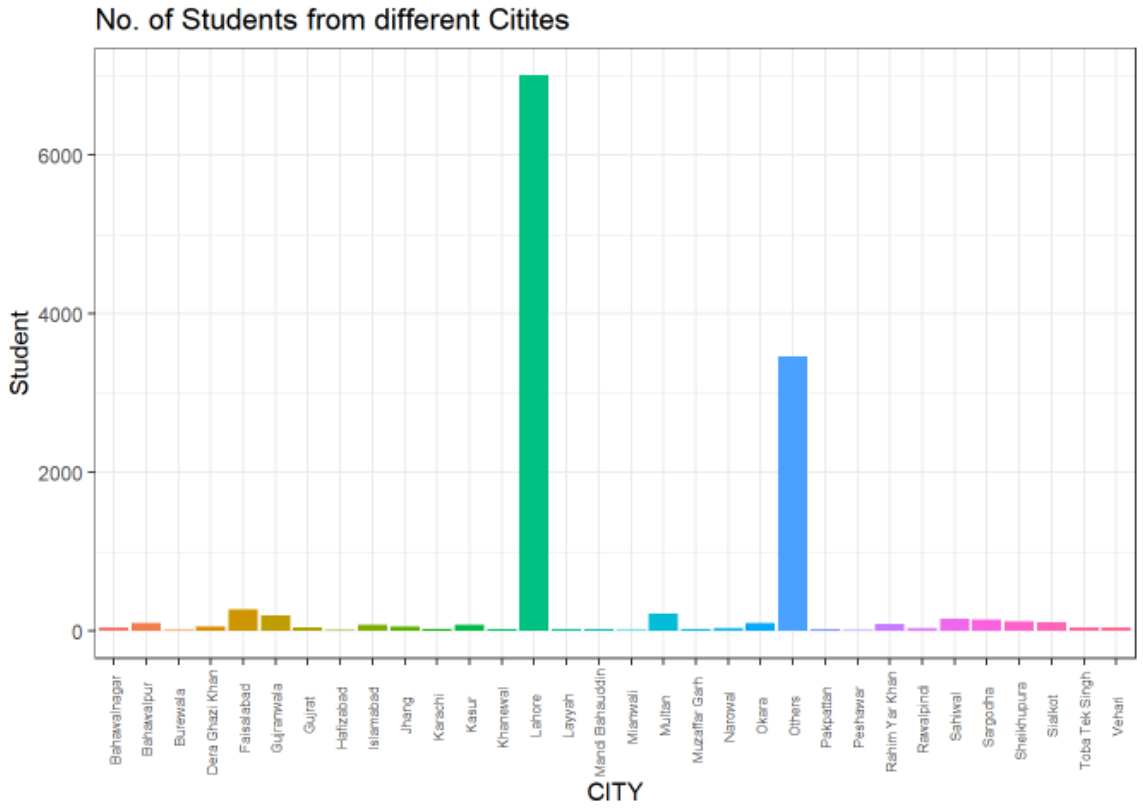


Figure 34. Number of Students from different cities - Lahore Campus

As obvious , the majority of students at Lahore Campus are from Lahore .

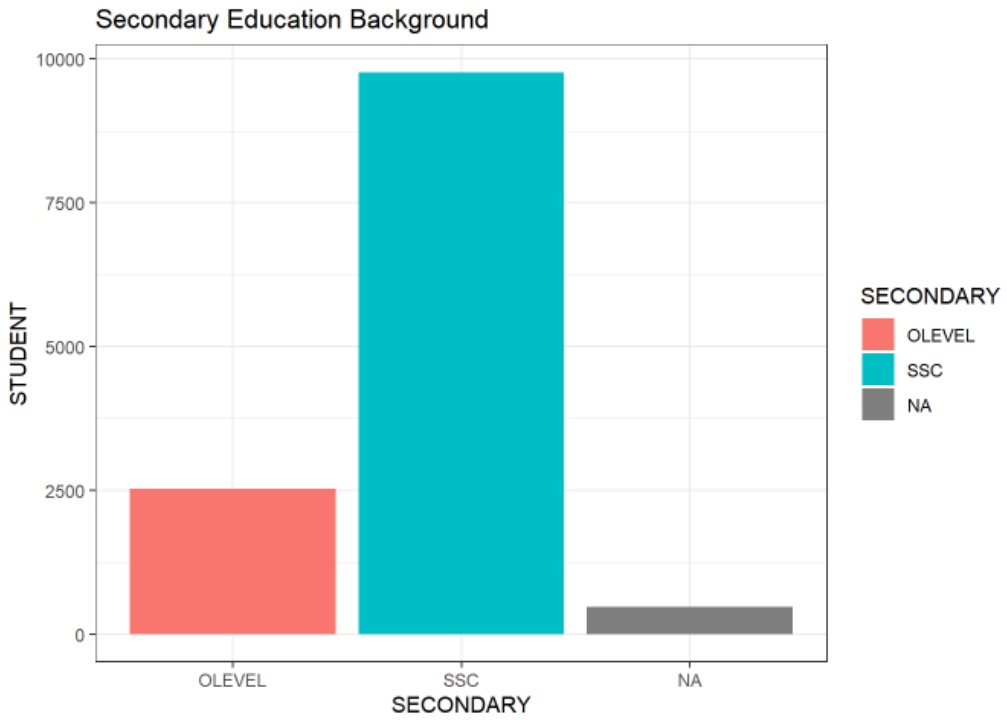


Figure 35. Secondary Education Background - Lahore Campus

At Lahore Campus , 76.58% students are from SSC background.

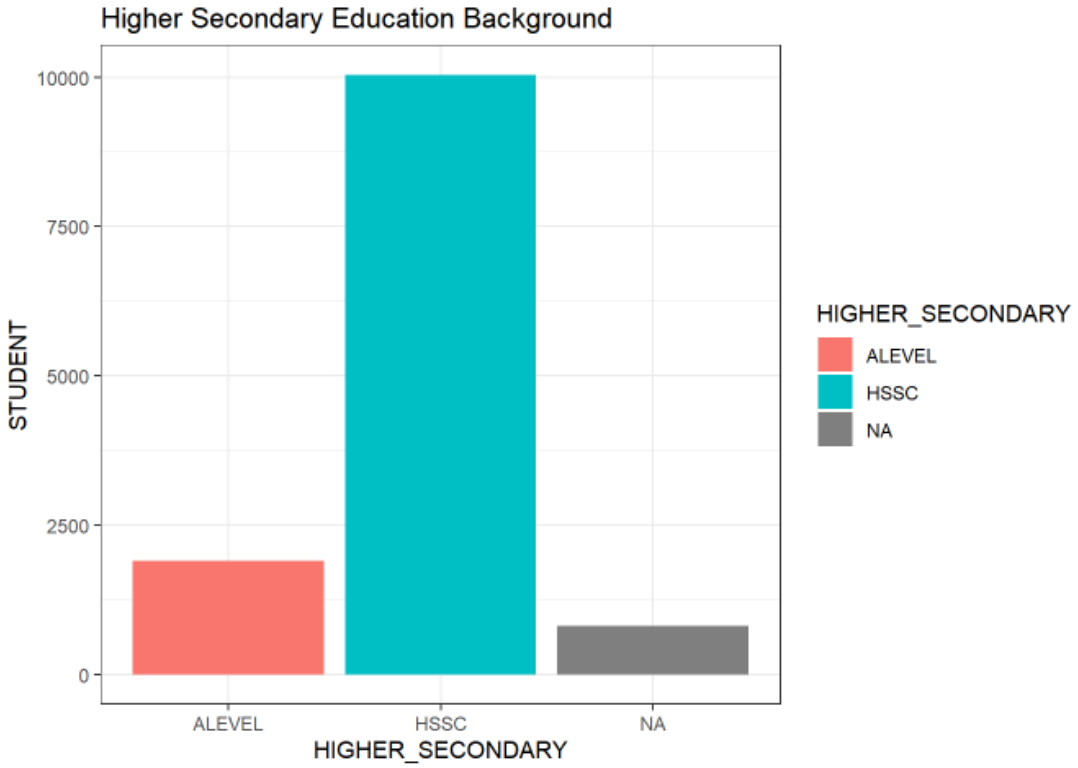


Figure 36. Higher Secondary Education Background - Lahore Campus

At Lahore Campus , 78.71% students are from HSSC background.

5.6 Visualization of the data of Peshawar Campus

Peshawar Campus offered 6 different degree programs .

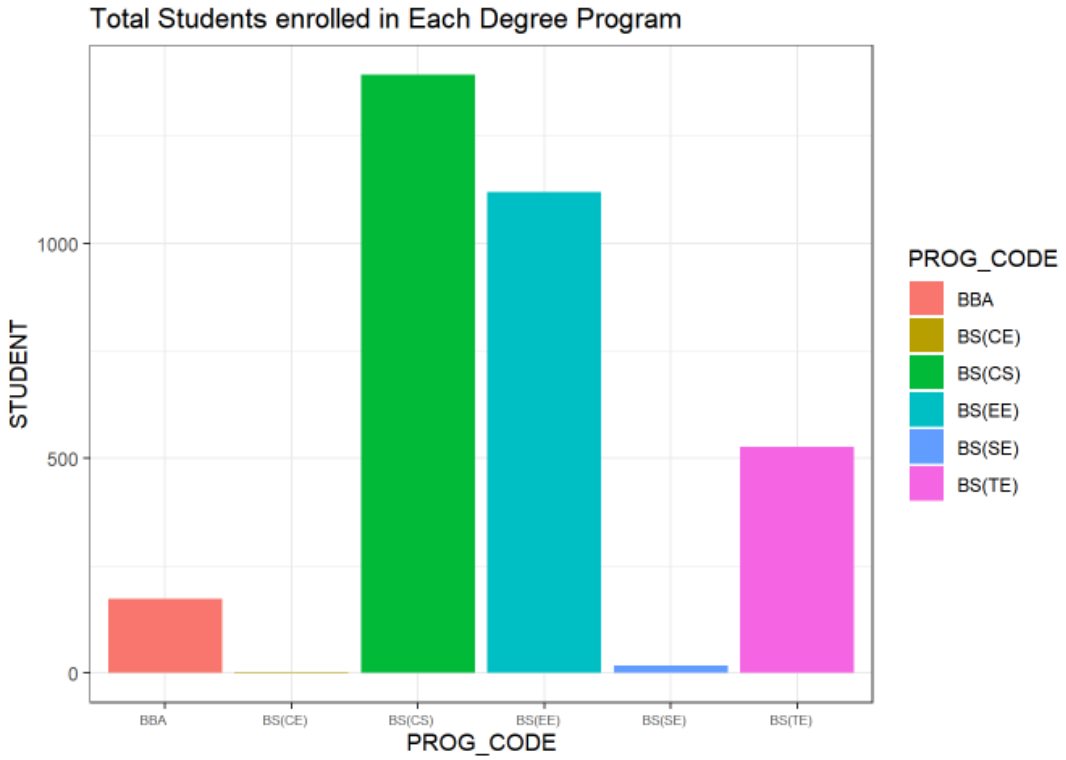


Figure 37. Total Students enrolled in each Degree Program - Peshawar Campus

At Peshawar Campus , a significant number of students are from BS(CS) and BS(EE). Also the count of students enrolled in BS(TE) is more than other cities.

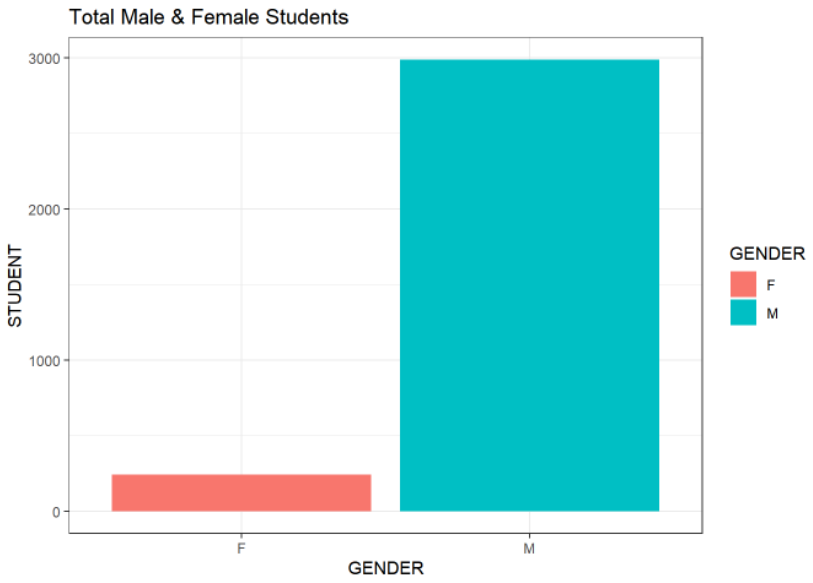


Figure 38. Total Male & Female Students - Peshawar Campus

Out of 3,229 students , Peshawar Campus has male students around 92.53% .

Students from 109 different cities took admission at Peshawar Campus. To visualize at a better level , only 44 cities are displayed after a minimum threshold of 10 students .

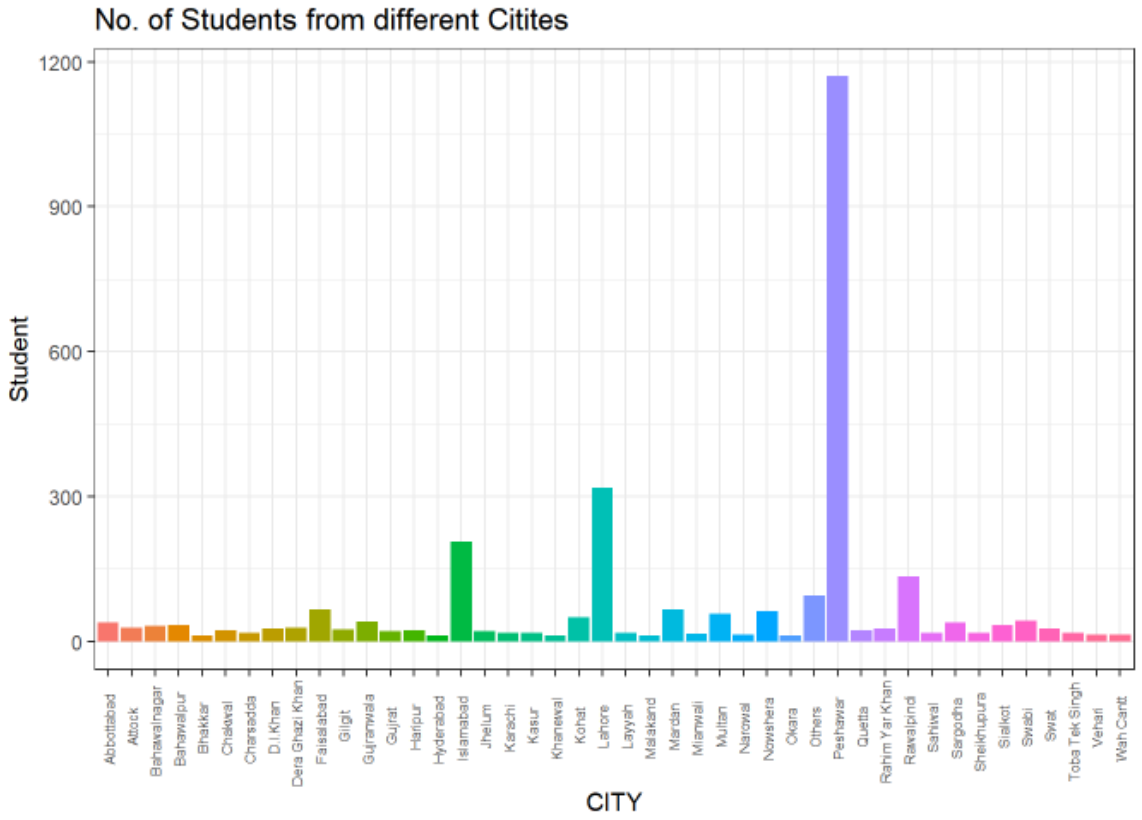


Figure 39. Number of Students from different cities - Peshawar Campus

In descending order students from Peshawar , Lahore and Hyderabad have majority representation.

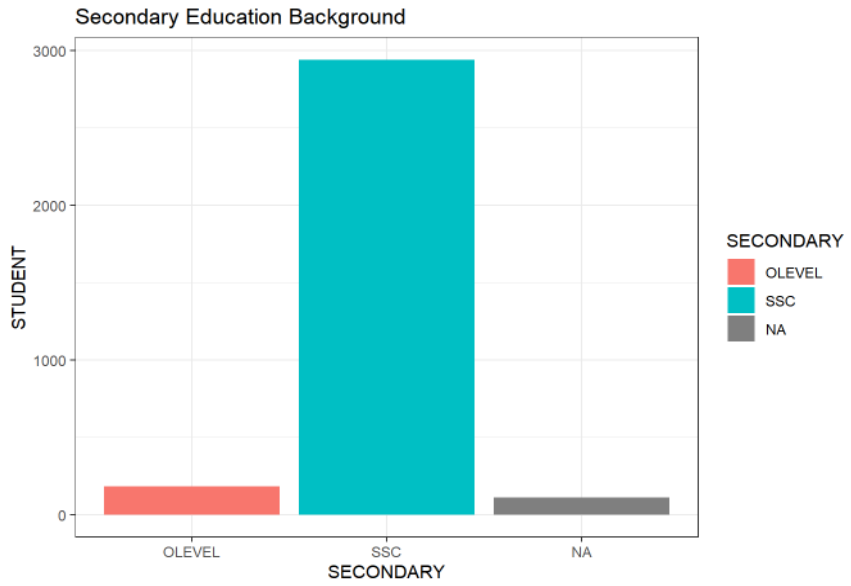


Figure 40. Secondary Education Background - Peshawar Campus

At Peshawar Campus , 91.04% students are from SSC background.

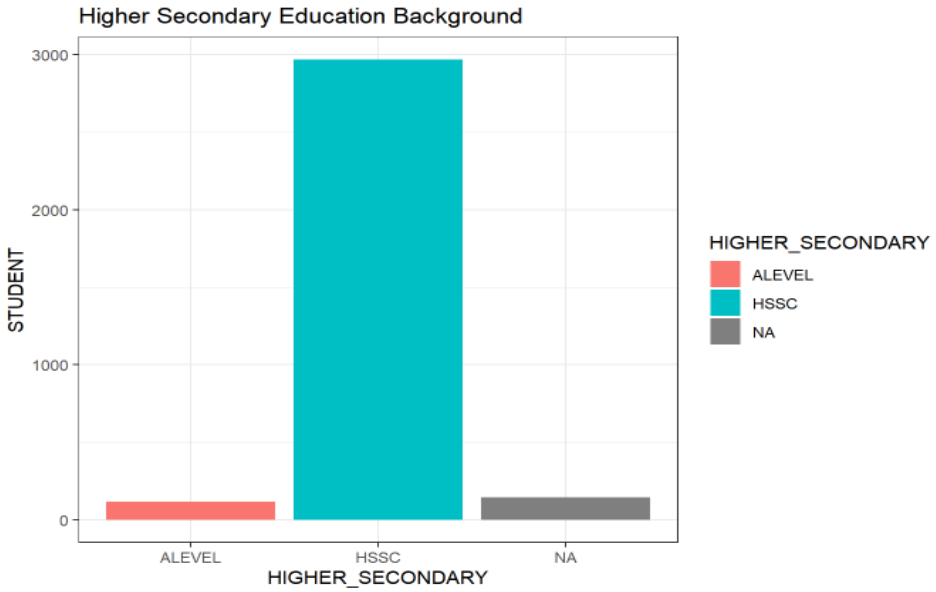


Figure 41. Higher Secondary Education Background - Peshawar Campus

At Peshawar Campus , 91.97% students are from HSSC background.

Chapter 6

Statistical Analysis

Now when the data is pre processed , transformed and cleaned , we move forward to the feature selection .

6.1 Not considering the data of warning students as a part of analysis

Since the warning students are outliers for our problem , initially analysis was done without the warning data. The results here are also explored in two ways: first for the whole data and then separately for each campus.

6.1.1 Working on the whole data

Question 1 : Do students from one campus perform better as compared to students from other campuses ? OR which campus has a higher average CGPA ?

Null Hypothesis 1: Campus affects CGPA
Alternative Hypothesis 1: Campus doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## CAMPUS         4    4.8  1.2124   6.157 6.29e-05 ***
## Residuals   2495  491.3  0.1969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 42. ANOVA - Campus vs CGPA

From Figure 42 , with p value less than significance value it can be seen that we reject the existence of any relationship between campus and CGPA .

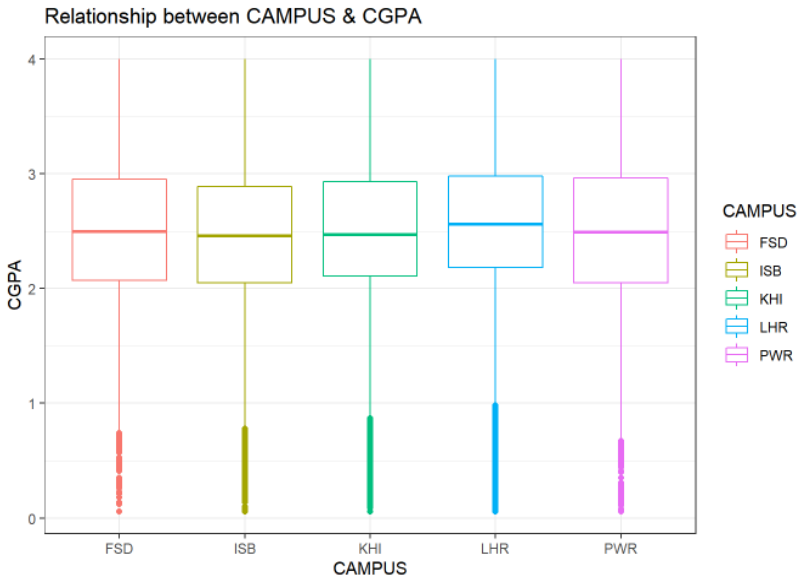


Figure 43. Box Plot - Campus vs CGPA

From Figure 43 also , it can be clearly seen that the mean CGPA for each campus is almost the same , so campus doesn't really matter in terms of student performance.

Question 2 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 2: Degree Program affects CGPA
Alternative Hypothesis 2: Degree Program doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## PROG_CODE      5   3.97  0.7938   4.015 0.00129 **
## Residuals   1194 236.05  0.1977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 44. ANOVA - Degree Program vs CGPA

From Figure 44 , with p value equal to significance value , we can state that yes students' performance varies for each degree program .

PROG_CODE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
BBA	2141	2.00	2.53	2.82	2.87	3.17	4.00
BS(AF)	581	2.01	2.41	2.69	2.77	3.08	3.97
BS(CE)	184	2.01	2.46	2.74	2.80	3.09	3.94
BS(CS)	6922	2.00	2.39	2.70	2.76	3.07	4.00
BS(CV)	286	2.10	2.59	2.86	2.90	3.21	3.94
BS(EE)	2335	2.00	2.45	2.77	2.82	3.15	3.99
BS(TE)	1696	2.00	2.51	2.81	2.86	3.17	4.00

Table 1 . Numerical Summary of Degree Programs for whole data

From Table 1 we can clearly see that students enrolled in BS(CV) perform better than the rest.

Question 3 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 3: Gender affects CGPA
Alternative Hypothesis 3: Gender doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## GENDER         1  20.62  20.625   104.8 <2e-16 ***
## Residuals   1198 235.76   0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 45. ANOVA - Gender vs CGPA

From the results of ANOVA , as p value is significantly less we can clearly see CGPA doesn't relate to gender. Both boys and girls perform equally regardless of gender .

Question 4 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 4: City affects CGPA
Alternative Hypothesis 4: City doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## CITY         17   9.85  0.5796   2.916 6.51e-05 ***
## Residuals  1062 211.11  0.1988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 46. ANOVA - City vs CGPA

From the results , a city has no impact on the student performance.

Question 5 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 5: Secondary education background affects CGPA
Alternative Hypothesis 5: Secondary education background doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## SECONDARY     1   10.2  10.227   48.22 4.44e-12 ***
## Residuals  3998  848.0   0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 47. ANOVA - Secondary Education vs CGPA

With p level less than significance value we can say that there is no difference between performance of students based on whether they belong to a matric background or an O levels background.

Question 6 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 6: School affects CGPA
Alternative Hypothesis 6: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## SCHOOL       36  11.87  0.3297   1.825 0.00257 **
## Residuals   703 127.02  0.1807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 48. ANOVA - School vs CGPA

Yes , depending on the school from which a student studied , his/her performance may vary .

SCHOOL	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
abbottabad	46	2.04	2.42	2.66	2.75	3.00	3.78
ajk board	43	2.05	2.38	2.57	2.76	3.15	3.81
akueb	41	2.13	2.40	2.73	2.81	3.10	3.94
bahawalpur	150	2.07	2.33	2.67	2.74	3.05	3.88
balochistan	31	2.08	2.58	2.78	2.88	3.20	3.74
beaconhouse garden town	100	2.06	2.54	2.96	2.93	3.26	3.90
bise	103	2.05	2.36	2.57	2.69	2.95	3.82
bise peshawar	21	2.04	2.54	2.77	2.72	2.98	3.25
blue horizon	71	2.06	2.55	2.76	2.80	3.09	3.72
cambridge international examinations	23	2.10	2.45	2.77	2.84	3.15	3.75
dera ghazi khan	25	2.11	2.49	2.66	2.71	2.85	3.86
dgkhan	75	2.02	2.29	2.54	2.57	2.76	3.49
excelsior	34	2.35	2.78	3.16	3.11	3.34	3.92
faisalabad	459	2.02	2.47	2.74	2.79	3.08	3.88
garrison acedemy for girls	22	2.20	2.48	2.73	2.73	2.91	3.85
gujranwala	229	2.02	2.33	2.63	2.72	3.07	3.94
hyderabad	94	2.05	2.29	2.70	2.72	3.10	3.98
karachi	745	2.00	2.33	2.61	2.69	2.95	3.96
lahor	1203	2.00	2.42	2.70	2.76	3.06	3.91
mirpurkhas	67	2.04	2.30	2.46	2.55	2.71	3.36
multan	155	2.00	2.33	2.58	2.65	2.90	3.99
other	2979	2.00	2.54	2.84	2.89	3.19	4.00
peshawar	145	2.03	2.35	2.63	2.71	2.97	3.84
private	64	2.01	2.47	2.70	2.76	3.08	3.77
rawalpindi	333	2.01	2.43	2.75	2.80	3.11	3.99
sacred heart cathederal school	31	2.07	2.50	2.84	2.77	3.09	3.38
saint marys academy	93	2.03	2.37	2.60	2.73	3.15	3.67
sargodha	120	2.04	2.39	2.67	2.73	2.92	3.87
st michaels convent	60	2.04	2.56	2.92	2.92	3.24	3.95
st pauls english high school	259	2.00	2.38	2.73	2.78	3.12	3.93
sukkur	44	2.09	2.34	2.51	2.63	2.85	3.72
the educators	24	2.13	2.43	2.64	2.74	3.06	3.84
university wensam college	1658	2.00	2.50	2.82	2.87	3.20	4.00
usman science school	2072	2.01	2.42	2.74	2.79	3.11	3.99

Table 2 . Numerical Summary of Schools for whole data

Table 2 shows that students from different schools perform differently , with students from Excelsior performing the best .

Question 7 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 7: Higher Secondary education background affects CGPA
Alternative Hypothesis 7: Higher Secondary education background doesn't affect CGPA .

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY    2    1.6  0.7977   3.865 0.0212 *
## Residuals        1497  309.0  0.2064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 49. ANOVA - Higher Secondary Education vs CGPA

With a p value greater than the significance value , it shows that there is a difference in performance depending on the higher secondary education background .

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	1843	2	2.53	2.87	2.90	3.24	4
HSSC	11758	2	2.43	2.73	2.79	3.10	4

Table 3 . Numerical Summary of Higher Secondary Education for whole data

We can see that students of A level, although being in the minority, have a better average than the students from HSC . So we can say that A level students are better performing than students from HSC .

Question 8 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 8: College affects CGPA
Alternative Hypothesis 8: College doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      29   9.64  0.3325   1.733 0.0108 *
## Residuals   570 109.39  0.1919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 50. ANOVA - College vs CGPA

Since p value is greater than significance value , we can conclude that college impacts a student's performance.

COLLEGE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
abbottabad	38	2.04	2.44	2.65	2.75	3.01	3.78
aisha bawany	217	2.05	2.47	2.76	2.81	3.10	3.82
bahria colleg karsaz habib rahmatullah road	37	2.13	2.28	2.70	2.72	2.96	3.71
balochistan	69	2.05	2.32	2.61	2.70	3.06	3.74
beaconhouse defence	21	2.34	2.48	2.90	2.90	3.24	3.78
british council peshawar	50	2.07	2.47	2.86	2.91	3.37	3.81
cambridg board	106	2.09	2.61	2.92	2.96	3.31	3.94
cambridge	699	2.00	2.51	2.82	2.87	3.20	3.99
commecs institite of business and emerging sciences	25	2.11	2.33	2.52	2.58	2.77	3.62
dera ismail khan	538	2.00	2.39	2.68	2.74	3.05	3.99
fbise	301	2.01	2.33	2.59	2.69	3.00	3.88
feder	776	2.00	2.44	2.76	2.80	3.11	3.94
feder board of intermedi and secondari educ	316	2.03	2.57	2.86	2.90	3.16	4.00

multan	105	2.00	2.34	2.64	2.70	3.05	3.64
others	3968	2.00	2.48	2.79	2.84	3.15	4.00
peshawar	83	2.03	2.34	2.61	2.72	3.00	3.89
private	62	2.08	2.40	2.82	2.86	3.20	3.98
punjab group of collges	1564	2.00	2.42	2.72	2.78	3.10	3.93
punjab technical	117	2.07	2.48	2.85	2.89	3.22	3.95
rawalpindi	368	2.01	2.40	2.70	2.76	3.09	3.95
sukkur	2787	2.00	2.42	2.70	2.76	3.06	3.98
summit education system	22	2.29	2.68	2.91	2.94	3.27	3.63
superior science higher secondary school	566	2.00	2.42	2.77	2.81	3.14	4.00
wensam college	297	2.00	2.53	2.84	2.90	3.24	4.00

Table 4 . Numerical Summary of Colleges for whole data

Based on the mean values from Table 4 , the students from British Council Peshawar , Beaconhouse Defence and Wensam college perform better than other students.

Question 9 : Does a semester of admission impact performance ?

Null Hypothesis 9: Admission Year affects CGPA
Alternative Hypothesis 9: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## FIRST_SEM    14   6.99  0.4995   2.485 0.00189 **
## Residuals   735 147.75  0.2010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 51. ANOVA - Admission Year vs CGPA

With an equal p value , yes the admission semester makes a difference in performance .

FIRST_SEM	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Fall 2001	446	2.04	2.57	2.85	2.91	3.22	4.00
Fall 2002	435	2.00	2.54	2.90	2.91	3.27	3.95
Fall 2003	477	2.01	2.49	2.78	2.85	3.18	4.00
Fall 2004	606	2.01	2.42	2.74	2.78	3.09	4.00
Fall 2005	644	2.00	2.45	2.78	2.82	3.14	3.96
Fall 2006	775	2.00	2.52	2.83	2.88	3.17	3.99
Fall 2007	817	2.00	2.55	2.83	2.90	3.23	4.00
Fall 2008	671	2.04	2.54	2.86	2.89	3.19	3.98
Fall 2009	866	2.01	2.53	2.88	2.87	3.18	3.98
Fall 2010	1042	2.01	2.40	2.70	2.78	3.10	3.98
Fall 2011	1181	2.00	2.42	2.71	2.77	3.09	3.93
Fall 2012	1639	2.00	2.35	2.63	2.71	3.01	4.00
Fall 2013	1608	2.00	2.37	2.64	2.73	3.05	3.99
Fall 2014	1554	2.01	2.42	2.70	2.76	3.06	3.99
Fall 2015	1201	2.00	2.54	2.82	2.87	3.18	3.96

Table 5. Numerical Summary of Semesters for whole data

The students taking admission in Fall 2001 and Fall 2002 performed better .

Question 10 : Does a graduating semester impact student performance ?

Null Hypothesis 10: Graduating Year affects CGPA
Alternative Hypothesis 10: Graduating year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM      34  91.81  2.7003   20.09 <2e-16 ***
## Residuals   1715 230.54  0.1344
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 52. ANOVA - Graduating Year vs CGPA

No , the semester of graduation has no impact on the student performance.

Question 11 : Are grades in school important for a better performance at university ?

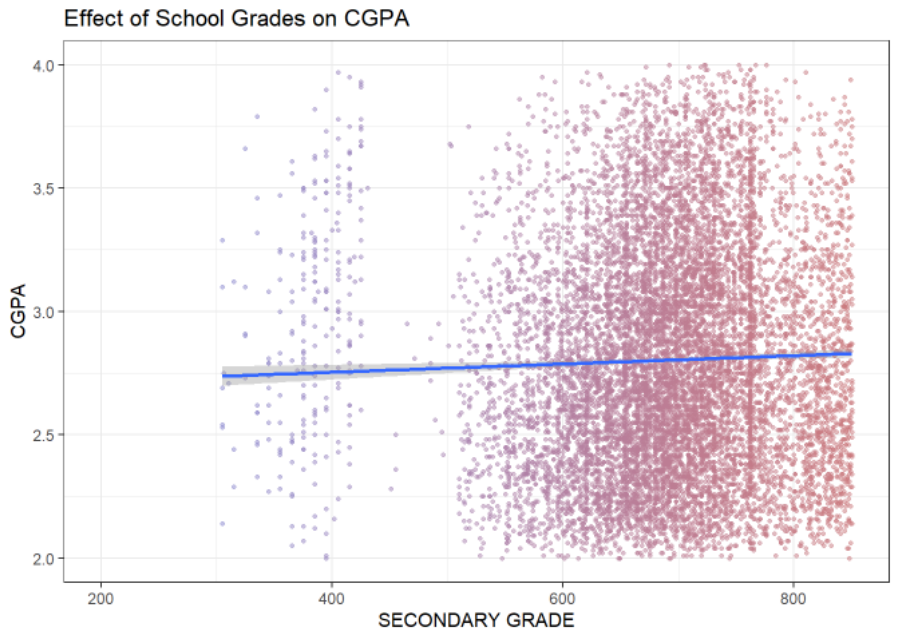


Figure 53. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  data$SEC_GRADE and data$CGPA
## t = 3.7297, df = 14143, p-value = 0.0001925
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.01487415 0.04780148
## sample estimates:
##        cor
## 0.03134632
```

Figure 54. Pearson Correlation - Secondary Grade vs CGPA

With a very low value of correlation it shows that there is no correlation between Secondary Grade and CGPA.

Question 12 : Are grades in college important for a better performance at university ?

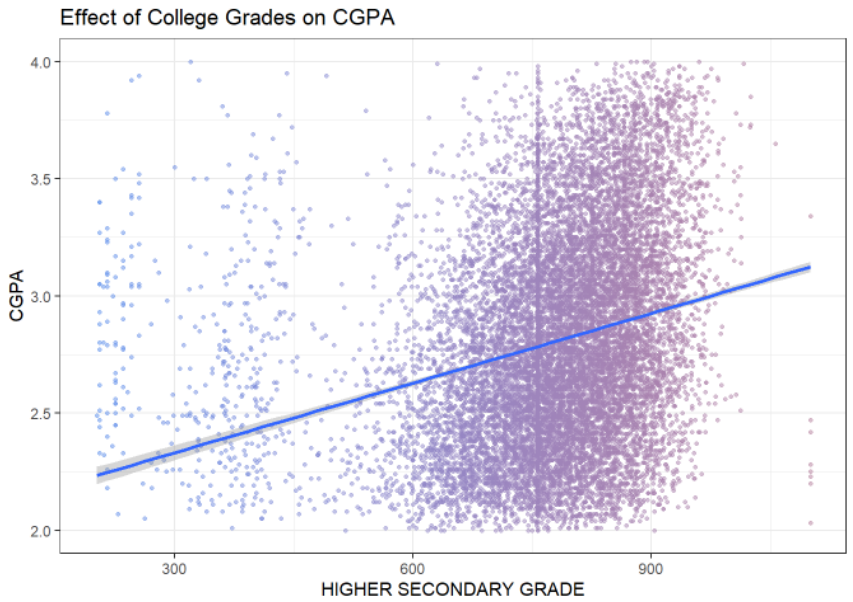


Figure 55. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  data$HIG_SEC_GRADE and data$CGPA
## t = 28.49, df = 14143, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2173245 0.2484958
## sample estimates:
##     cor
## 0.23297
```

Figure 56. Pearson Correlation - Higher Secondary Grade vs CGPA

Yes , the grades in college have an impact on a student's performance in university.

Question 13 : What are the important core CS courses that affect a student's performance ?

```
##
## Recursive feature selection
##
## Outer resampling method: Cross-Validated (2 fold, repeated 2 times)
##
## Resampling performance over subset size:
##
##  Variables   RMSE Rsquared    MAE    RMSESD RsquaredSD     MAESD Selected
##          1 0.3558   0.4244 0.2841 0.0057587   0.017324 0.0053070
##          2 0.2987   0.5961 0.2384 0.0035251   0.009528 0.0025645
##          3 0.2670   0.6787 0.2136 0.0005757   0.001489 0.0007708
##          4 0.2476   0.7245 0.1983 0.0004280   0.002751 0.0004313
##          5 0.2331   0.7568 0.1867 0.0004413   0.001471 0.0007179
##          6 0.2251   0.7697 0.1788 0.0007954   0.002569 0.0003316
##         11 0.2036   0.8135 0.1622 0.0010656   0.001538 0.0012948        *
##
## The top 5 variables (out of 11):
##    probability and statistics, database systems, operating systems, theory of automata, linear algebra
```

Figure 57. Recursive Feature Elimination - CS Courses

From the results of RFE , 5 important courses that are important for a CS degree are Probability and statistics , Database Systems , Operating Systems , Theory of Automata and Linear Algebra .

```
##  [1] "`database systems`"           "`theory of automata`"
##  [3] "`probability and statistics`" "`operating systems`"
##  [5] "`linear algebra`"             "`computer programming`"
##  [7] "`discrete structures`"        "`introduction to computing`"
##  [9] "`calculus - ii`"              "`data structures`"
## [11] "`calculus - i`"
```

Figure 58. Stepwise Backward and Forward Selection - CS Courses

The stepwise backward and forward selection approach also supports the results of RFE . Although the order of ranking is different , the courses are the same .

Question 14 : What are the important core EE courses that affect a student's performance ?

```
##
## Recursive feature selection
##
## Outer resampling method: Cross-Validated (2 fold, repeated 2 times)
##
## Resampling performance over subset size:
##
##  Variables   RMSE Rsquared    MAE   RMSESD RsquaredSD     MAESD Selected
##          1 0.3466   0.4404 0.2783 0.004208   0.018812 0.0025022
##          2 0.2923   0.6028 0.2334 0.006082   0.014075 0.0058363
##          3 0.2635   0.6784 0.2103 0.002489   0.010439 0.0028003
##          4 0.2393   0.7364 0.1902 0.002552   0.004500 0.0028044
##          5 0.2239   0.7704 0.1785 0.001372   0.005690 0.0014982
##         13 0.1779   0.8570 0.1421 0.001008   0.005382 0.0001621        *
##
## The top 5 variables (out of 13):
##    µp interfacing and programming, signals and systems, probability and random processes, electromagnetic theory, electro
mechanical systems
```

Figure 59. Recursive Feature Elimination - EE Courses

From the results of RFE , 5 important courses that are important for an EE degree are Interfacing and Programming , Signals and Systems , Probability and Random Processes , Electromagnetic theory and Mechanical Systems .

```
##  [1] "`signals and systems`"
##  [2] "`µp interfacing and programming`"
##  [3] "`linear algebra`"
##  [4] "`electromagnetic theory`"
##  [5] "`electro mechanical systems lab`"
##  [6] "`probability and random processes`"
##  [7] "`applied calculus`"
##  [8] "`electro mechanical systems`"
##  [9] "`signals and systems lab`"
## [10] "`complex variables and transforms `"
## [11] "`µp interfacing and programming lab`"
## [12] "`engineering workshop`"
## [13] "`differential equations`"
```

Figure 60. Stepwise Backward and Forward Selection - EE Courses

However , stepwise backward and forward selection , selects Linear Algebra and Electro mechanical systems labs as also important courses of a EE degree .

Question 15 : What are the important core BBA courses that affect a student's performance ?

```
##
## Recursive feature selection
##
## Outer resampling method: Cross-Validated (2 fold, repeated 2 times)
##
## Resampling performance over subset size:
##
##  Variables   RMSE Rsquared    MAE   RMSESD RsquaredSD    MAESD Selected
##          1 0.3866   0.3597 0.3125 0.008240   0.035434 0.009139
##          2 0.3122   0.5843 0.2483 0.003224   0.014994 0.002528
##          3 0.2829   0.6602 0.2267 0.005055   0.008174 0.005052
##          4 0.2632   0.7075 0.2090 0.005514   0.011934 0.004414
##          5 0.2576   0.7215 0.2045 0.005616   0.008036 0.004791
##          6 0.2512   0.7304 0.1973 0.004167   0.006537 0.003669        *
##
## The top 5 variables (out of 6):
##    business statistics, fundamentals of management, business communication - i, financial accounting, fundamentals of acc
ounting
```

Figure 61. Recursive Feature Eliminations - BBA Courses

According to RFE , for a BBA degree the top 5 important courses are Business Statistics , Fundamentals of Management , Business Communication , Financial Accounting and Fundamentals of Accounting .

```
## [1] "`fundamentals of management`" "`business statistics`"
## [3] "`business communication - i`" "`financial accounting`"
## [5] "`it in business`"             "`fundamentals of accounting`"
```

Figure 62. Recursive Feature Eliminations - CS Courses

Also if we try stepwise backward and forward selection , it also shows almost the same result with an addition of IT in business .

6.1.2 Working on the data of Faisalabad Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affects CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## PROG_CODE     3   1.28  0.4277   2.454 0.0622 .
## Residuals   661 115.21  0.1743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 63. ANOVA - Degree Program vs CGPA

With p value greater than significance value , we can state that yes students' performance varies for each degree program .

PROG_CODE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
BBA	83	2.14	2.39	2.69	2.79	3.12	3.84
BS(AF)	39	2.02	2.45	2.79	2.81	3.12	3.93
BS(CS)	495	2.05	2.45	2.72	2.77	3.06	4.00
BS(EE)	48	2.26	2.63	2.92	2.94	3.21	3.72

Table 6. Numerical Summary of Degree Programs - Faisalabad Campus

From Table 6 we can clearly see that in the Faisalabad campus students enrolled in BS(EE) perform better than the rest .

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affects CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## GENDER        1   2.54  2.5448   14.83 0.000158 ***
## Residuals   198  33.97  0.1716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 64. ANOVA - Gender vs CGPA

From the results of ANOVA , as p value is almost equal to significance value we can clearly see CGPA relates to gender.

GENDER	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
F	112	2.02	2.62	2.96	2.94	3.23	3.93
M	553	2.05	2.43	2.69	2.75	3.05	4.00

Table 7. Numerical Summary of Gender - Faisalabad Campus

Girls perform better than boys in terms of CGPA .

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affects CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## CITY         12   5.43  0.4521   2.701 0.00149 **
## Residuals   575  96.26  0.1674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 65. ANOVA - City vs CGPA

Since p value and significance value are equal we go with the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Bahawalpur	7	2.20	2.68	2.82	2.77	2.95	3.10
Chiniot	42	2.16	2.51	2.74	2.84	3.19	3.84
Faisalabad	345	2.02	2.50	2.83	2.86	3.19	4.00
Gujranwala	9	2.26	2.40	2.73	2.73	2.94	3.20
Lahore	100	2.05	2.38	2.62	2.66	2.98	3.42
Multan	6	2.21	2.31	2.48	2.47	2.65	2.69
Others	7	2.07	2.33	2.55	2.59	2.83	3.24
Sahiwal	9	2.16	2.57	2.62	2.72	3.13	3.28
Sargodha	21	2.19	2.52	2.74	2.67	2.77	3.21
Sheikhupura	11	2.26	2.70	2.78	2.87	3.11	3.30
Sialkot	9	2.18	2.25	2.52	2.74	3.20	3.45
Toba Tek Singh	15	2.07	2.44	2.67	2.61	2.80	3.10
Vehari	7	2.25	2.31	2.36	2.65	2.91	3.49

Table 8. Numerical Summary of Cities - Faisalabad Campus

Students from Sheikhupura and Faisalabad cities perform better in the Faisalabad Campus .

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary education background affects CGPA
Alternative Hypothesis 4: Secondary education background doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY     1  0.719  0.7185    3.48 0.0644 .
## Residuals   126 26.012  0.2064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 66. ANOVA - Secondary Education vs CGPA

With p value greater than significance value we can say that there is a difference between performance of students based on whether they belong to a matric background or an O levels background.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	64	2.05	2.61	2.92	2.95	3.29	4.00
SSC	601	2.02	2.43	2.70	2.76	3.06	3.88

Table 9. Numerical Summary of Secondary Education - Faisalabad Campus

The students from O levels in Faisalabad Campus perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affects CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL      10  1.619  0.1619   1.146  0.337
## Residuals   99 13.994  0.1414
```

Figure 67. ANOVA - School vs CGPA

Yes , depending on the school from which a student studied , his/her performance may vary .

COLLEGE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
aisha bawany	93	2.11	2.62	2.91	2.89	3.12	3.72
balochistan	17	2.12	2.30	2.34	2.57	2.69	3.49
cambridge	28	2.05	2.78	3.19	3.12	3.44	3.93
dera ismail khan	58	2.06	2.45	2.75	2.75	2.94	3.76
khanewal public	22	2.12	2.36	2.67	2.66	2.90	3.47
others	16	2.25	2.32	2.59	2.65	2.87	3.51
rawalpindi	20	2.06	2.44	2.67	2.75	3.18	3.45
sukkur	365	2.02	2.43	2.69	2.75	3.05	3.88

Table 10. Numerical Summary of Schools - Faisalabad Campus

Table 10 shows that students from different schools perform differently , with students from Aisha Bawany performing the best .

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary education background affects CGPA
Alternative Hypothesis 6: Higher Secondary education background doesn't affect CGPA .

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY  1  0.888  0.8883   4.476 0.0376 *
## Residuals        78 15.478  0.1984
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 68. ANOVA - Higher Secondary Education vs CGPA

With a p value greater than the significance value , it shows that there is a difference in performance depending on the higher secondary education background .

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	48	2.05	2.72	3.08	3.01	3.32	4.00
HSSC	615	2.02	2.44	2.70	2.76	3.06	3.88

Table 11. Numerical Summary of Higher Secondary Education - Faisalabad Campus

From Table 11 , we can see that students of A level, although being in the minority, have a better average than the students from HSC . So we can say that A level students are better performing than students from HSC .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affects CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      29   9.64  0.3325   1.733 0.0108 *
## Residuals   570 109.39  0.1919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 69. ANOVA - College vs CGPA

Since p value is greater than significance value , we can conclude that college impacts a student's performance.

SCHOOL	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
bahawalpur	21	2.12	2.34	2.68	2.68	2.93	3.51
bise	39	2.13	2.39	2.68	2.76	3.08	3.82
blue horizon	28	2.06	2.56	2.87	2.82	3.05	3.45
faisalabad	302	2.02	2.51	2.77	2.83	3.13	3.88
feder	13	2.14	2.38	2.62	2.57	2.77	2.89
gujranwala	30	2.06	2.37	2.64	2.71	3.17	3.53
lahor	67	2.05	2.37	2.62	2.65	2.92	3.32
multan	17	2.12	2.25	2.32	2.52	2.63	3.49
other	20	2.14	2.28	2.56	2.64	3.00	3.43
sargodha	19	2.19	2.40	2.73	2.65	2.80	3.21
university wensam college	64	2.05	2.66	3.00	2.99	3.36	4.00

Table 12. Numerical Summary of Colleges - Faisalabad Campus

Based on the mean values from Table 12 , the students from University Wensam College , perform better than other students.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affects CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM     3   3.17  1.0554   6.194 0.000375 ***
## Residuals   653 111.27  0.1704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 70. ANOVA - Admission Year vs CGPA

With an almost equal p value , yes the admission semester makes a difference in performance .

FIRST_SEM	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Fall 2012	151	2.05	2.47	2.69	2.78	3.08	4.00
Fall 2013	203	2.02	2.37	2.65	2.70	3.00	3.84
Fall 2014	169	2.06	2.45	2.76	2.81	3.18	3.93
Fall 2015	134	2.11	2.62	2.89	2.90	3.16	3.76

Table 13. Numerical Summary of Semesters - Faisalabad Campus

The students taking admission in Fall 2015 performed the best .

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduating Year affects CGPA
Alternative Hypothesis 9: Graduating year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## LAST_SEM      6   9.22  1.5369   9.493 5.53e-10 ***
## Residuals   607  98.27  0.1619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 71. ANOVA - Graduating Year vs CGPA

No , the semester of graduation has no impact on the student performance.

Question 10 : Are grades in school important for a better performance at university ?

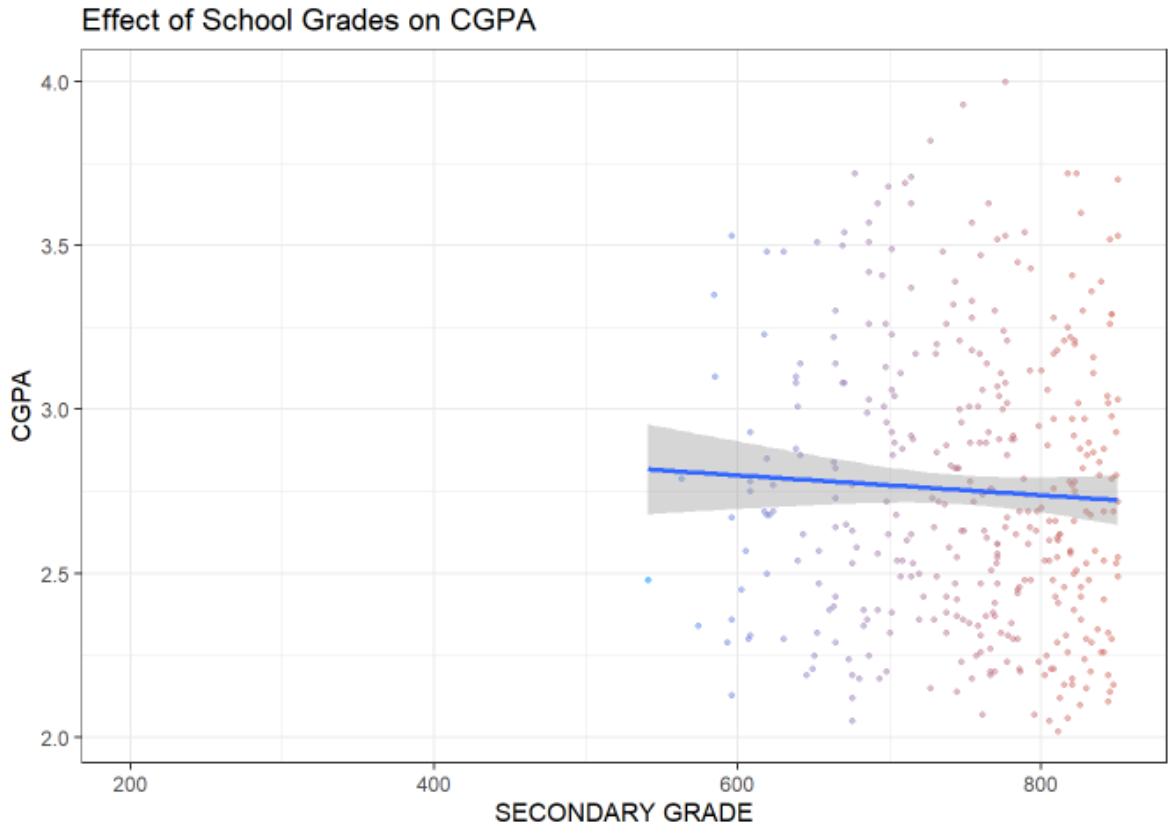


Figure 72. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  FAISALABAD_data$SEC_GRADE and FAISALABAD_data$CGPA
## t = 1.5822, df = 663, p-value = 0.1141
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.01476781  0.13672185
## sample estimates:
##       cor
## 0.06133021
```

Figure 73. Pearson Correlation - Secondary Grade vs CGPA

With a very low value of correlation it shows that there is no correlation between Secondary Grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

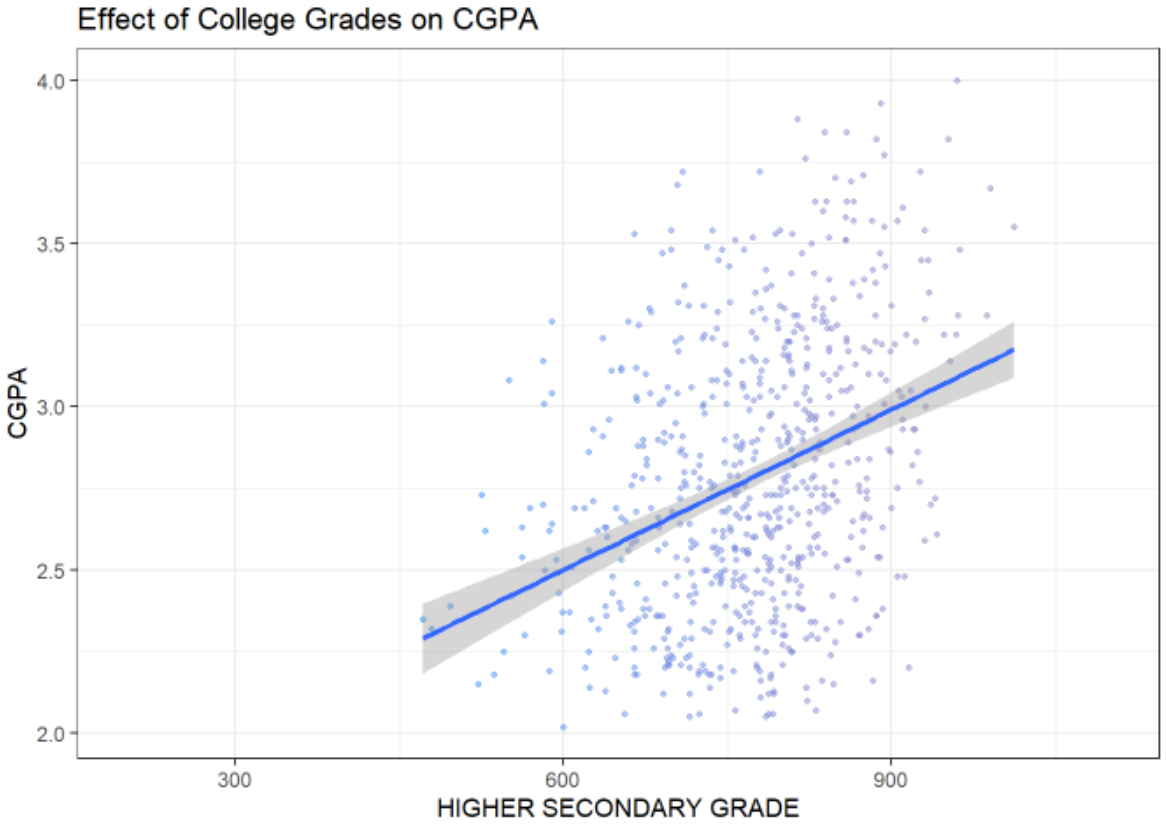


Figure 74. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  FAISALABAD_data$HIG_SEC_GRADE and FAISALABAD_data$CGPA
## t = 7.048, df = 663, p-value = 4.564e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1918323 0.3333489
## sample estimates:
##       cor
## 0.2640109
```

Figure 75. Pearson Correlation - Higher Secondary Grade vs CGPA

Yes , the grades in college have an impact on a student's performance in university.

6.1.3 Working on the data of Islamabad Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affects CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## PROG_CODE      4   13.9   3.471   17.14 6.17e-14 ***
## Residuals   3776  765.0   0.203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 76. ANOVA - Degree Program vs CGPA

With p value smaller than significance value , we can state that no students' performance has no correlation with degree program .

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affects CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## GENDER         1   20.9  20.914   103.9 <2e-16 ***
## Residuals   1998  402.1   0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 77. ANOVA - Gender vs CGPA

From the results of ANOVA , as p value is way smaller than significance value we can clearly see CGPA is not affected by gender.

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affects CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## CITY          30   15.8  0.5277   2.585 5.24e-06 ***
## Residuals   3550  724.6  0.2041
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 78. ANOVA - City vs CGPA

Since p value is smaller than significance value we go with the alternative hypothesis.

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary education background affects CGPA
Alternative Hypothesis 4: Secondary education background doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY     1   2.31  2.3061   10.87  0.001 **
## Residuals  1398 296.56  0.2121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 79. ANOVA - Secondary Education vs CGPA

With p value equal to significance value we can say that there is a difference between performance of students based on whether they belong to a matric background or an O levels background.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	706	2	2.53	2.84	2.87	3.2	3.95
SSC	3015	2	2.44	2.73	2.79	3.1	4.00

Table 14. Numerical Summary of Secondary Education - Islamabad Campus

The students from O levels in Islamabad Campus perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affects CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL       26   40.2  1.5467   7.911 <2e-16 ***
## Residuals  3662  715.9  0.1955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 80. ANOVA - School vs CGPA

As p value is very small than significance value , we reject the null hypothesis .\

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary education background affects CGPA
Alternative Hypothesis 6: Higher Secondary education background doesn't affect CGPA .

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## HIGHER_SECONDARY    1   4.11   4.111   19.94 8.65e-06 ***
## Residuals        1398 288.25   0.206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 81. ANOVA - Higher Secondary Education vs CGPA

With a p value smaller than the significance value , it shows that there is no difference in performance depending on the higher secondary education background .

.

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affects CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      18   33.2  1.8467   9.311 <2e-16 ***
## Residuals  3708  735.4  0.1983
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 82. ANOVA - College vs CGPA

Since p value is smaller than significance value , we can conclude that college doesn't impact a student's performance.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affects CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## FIRST_SEM    12   44.4   3.699   19.32 <2e-16 ***
## Residuals  3544  678.6   0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 83. ANOVA - Admission Year vs CGPA

The alternative hypothesis is accepted , as p value is less than significance value . .

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduating Year affects CGPA
Alternative Hypothesis 9: Graduating year doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM    19  64.44   3.392   20.45 <2e-16 ***
## Residuals  980 162.55   0.166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 84. ANOVA - Graduating Year vs CGPA

No , the semester of graduation has no impact on the student performance.

Question 10 : Are grades in school important for a better performance at university ?

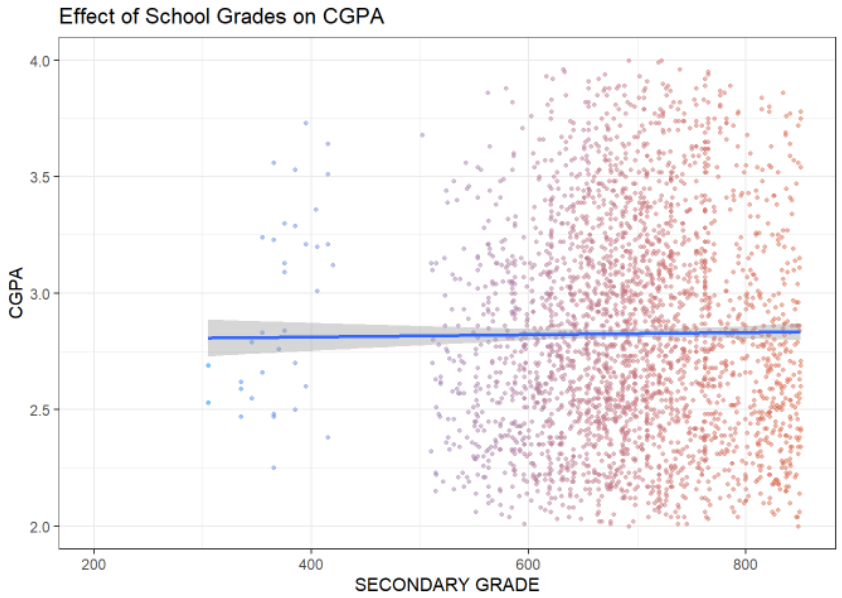


Figure 85. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  ISLAMABAD_data$SEC_GRADE and ISLAMABAD_data$CGPA
## t = -2.3964, df = 3779, p-value = 0.01661
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.070741698 -0.007085435
## sample estimates:
##         cor
## -0.03895309
```

Figure 86. Pearson Correlation - Secondary Grade vs CGPA

For Islamabad Campus , there is a negative impact of good school grades on CGPA at university .

Question 11 : Are grades in college important for a better performance at university ?

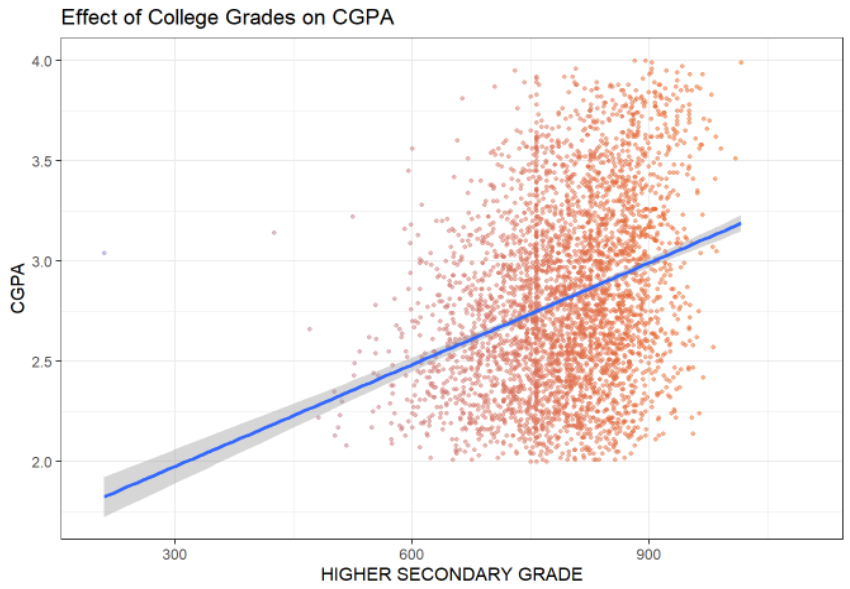


Figure 87. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  ISLAMABAD_data$HIG_SEC_GRADE and ISLAMABAD_data$CGPA
## t = 19.314, df = 3779, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2704512 0.3284814
## sample estimates:
##       cor
## 0.2997435
```

Figure 88. Pearson Correlation - Higher Secondary Grade vs CGPA

Yes , the grades in college have an impact on a student's performance in university.

6.1.4 Working on the data of Karachi Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affects CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## PROG_CODE     3  11.84   3.948   18.45 1.4e-11 ***
## Residuals   796 170.34   0.214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 89. ANOVA - Degree Program vs CGPA

With p value smaller than significance value , we can state that no students' performance does not vary for each degree program .

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affects CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## GENDER         1  15.34  15.341   72.33 <2e-16 ***
## Residuals   1198 254.10   0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 90. ANOVA - Gender vs CGPA

From the results of ANOVA , as p value is smaller than significance value we can clearly see CGPA does not relate to gender.

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affects CGPA

Alternative Hypothesis 3: City doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## CITY         69   18.0  0.2603   1.189  0.137
## Residuals  3872  847.6  0.2189
```

Figure 91. ANOVA - City vs CGPA

Since p value and significance value are equal we go with the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Badin	11	2.12	2.28	2.56	2.72	3.06	3.60
Dadu	23	2.05	2.29	2.47	2.64	2.88	3.63
Ghotki	17	2.05	2.24	2.47	2.56	2.61	3.83
Hyderabad	64	2.08	2.37	2.75	2.80	3.15	3.98
Islamabad	25	2.01	2.45	2.85	2.83	3.13	3.76
Karachi	2686	2.00	2.34	2.63	2.72	3.03	4.00
Khairpur	21	2.09	2.26	2.52	2.58	2.88	3.25
Lahore	27	2.16	2.33	2.76	2.88	3.39	3.91
Larkana	11	2.08	2.54	3.05	3.00	3.31	3.93
Multan	12	2.18	2.32	2.42	2.64	3.02	3.39
Others	807	2.00	2.34	2.65	2.74	3.06	3.96
Quetta	30	2.29	2.47	2.80	2.95	3.47	3.77
Rahim Yar Khan	20	2.10	2.44	2.62	2.70	2.96	3.90
Rawalpindi	10	2.12	2.29	2.77	2.80	3.29	3.64
Umerkot	14	2.23	2.38	2.66	2.66	2.95	3.07

Table 15. Numerical Summary of Cities - Karachi Campus

In the Karachi campus students from Quetta outperform others. Also students from Lahore and Islamabad perform better in the Karachi Campus .

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary education background affects CGPA
Alternative Hypothesis 4: Secondary education background doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY     1   1.39  1.3853   6.141 0.0134 *
## Residuals   998 225.14  0.2256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 92. ANOVA - Secondary Education vs CGPA

With p value greater than significance value we can say that there is a difference between performance of students based on whether they belong to a matric background or an O levels background.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	510	2	2.38	2.72	2.79	3.14	3.99
SSC	3388	2	2.34	2.63	2.72	3.04	4.00

Table 16. Numerical Summary of Secondary Education - Karachi Campus

The students from O levels in Karachi Campus perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affects CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## SCHOOL       27   11.0  0.4087   1.871 0.00414 **
## Residuals  3653  797.8  0.2184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 93. ANOVA - School vs CGPA

Yes , depending on the school from which a student studied , his/her performance may vary .

SCHOOL	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ajk board	27	2.16	2.38	2.61	2.83	3.38	3.80
akueb	36	2.13	2.44	2.78	2.84	3.11	3.94
bahawalpur	14	2.10	2.47	2.75	2.77	3.02	3.88
beaconhouse garden town	33	2.06	2.39	2.91	2.81	3.11	3.63
bvs parsi high school abdullah haroon road	8	2.31	2.73	3.00	3.08	3.29	4.00
cambridge international examinations	16	2.10	2.40	2.81	2.89	3.44	3.75
chiniot islamia	17	2.04	2.27	2.46	2.54	2.65	3.80
feder	75	2.07	2.41	2.62	2.76	3.07	3.91
gulistan shah abdul latif	9	2.04	2.14	2.53	2.60	3.05	3.16
sacred heart cathederal school	18	2.07	2.31	2.52	2.59	2.82	3.26
saint marys academy	69	2.06	2.35	2.54	2.72	3.08	3.67
st michaels convent	35	2.04	2.46	2.81	2.84	3.16	3.95
st pauls english high school	167	2.00	2.34	2.60	2.71	3.02	3.93
sukkur	40	2.09	2.31	2.51	2.62	2.85	3.72
sultan mohammad shah aga khan	11	2.17	2.52	2.86	2.96	3.33	3.95
the educators	10	2.14	2.46	2.64	2.82	3.26	3.84

Table 17 Numerical Summary of Schools- Karachi Campus

Table 17 shows that students from different schools perform differently , with students from BVS Parsi high school performing the best .

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary education background affects CGPA
Alternative Hypothesis 6: Higher Secondary education background doesn't affect CGPA .

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY   1   1.75  1.7483   8.009 0.0048 **
## Residuals        638 139.26  0.2183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 94. ANOVA - Higher Secondary Education vs CGPA

With a p value greater than the significance value , it shows that there is a difference in performance depending on the higher secondary education background .

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	324	2	2.43	2.78	2.83	3.16	3.99
HSSC	3256	2	2.33	2.61	2.71	3.02	4.00

Table 18. Numerical Summary of Higher Secondary Education- Karachi Campus

We can see that students of A level, although being in the minority, have a better average than the students from HSC . So we can say that A level students are better performing than students from HSC .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affects CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## COLLEGE       18   14.2   0.790   3.657 2.68e-07 ***
## Residuals   3678  794.5   0.216
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 95. ANOVA - College vs CGPA

Since p value is smaller than significance value , we can conclude that college does not impact a student's performance.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affects CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM     14   17.6  1.2558    5.81 2.17e-11 ***
## Residuals   3868  836.0  0.2161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 96. ANOVA - Admission Year vs CGPA

With a smaller p value , no, the admission semester does not make a difference in performance .

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduating Year affects CGPA
Alternative Hypothesis 9: Graduating year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM      17   54.7   3.215   16.01 <2e-16 ***
## Residuals   3277  658.0   0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 97. ANOVA - Graduation Year vs CGPA

No , the semester of graduation has no impact on the student performance.

Question 10 : Are grades in school important for a better performance at university ?

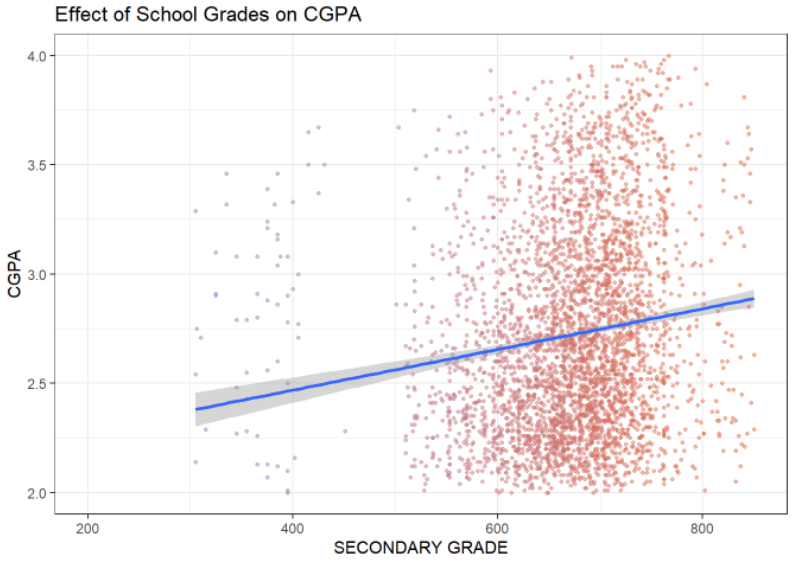


Figure 98. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  KARACHI_data$SEC_GRADE and KARACHI_data$CGPA
## t = 11.399, df = 3940, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1482856 0.2087314
## sample estimates:
##       cor
## 0.1786771
```

Figure 99. Pearson Correlation - Secondary Grade vs CGPA

With a very low value of correlation it shows that there is a weak correlation between Secondary Grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

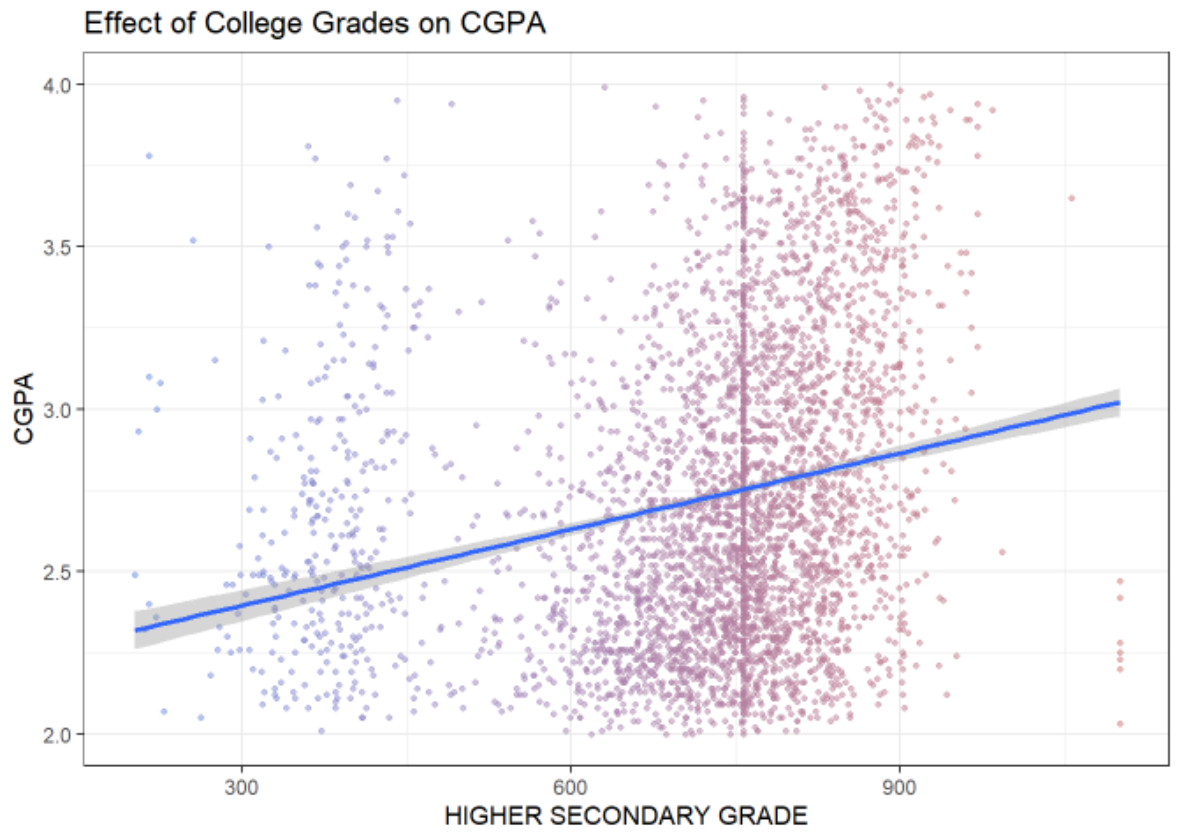


Figure 100. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  KARACHI_data$HIG_SEC_GRADE and KARACHI_data$CGPA
## t = 13.3, df = 3940, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1772196 0.2369765
## sample estimates:
##       cor
## 0.2072914
```

Figure 101. Pearson Correlation - Higher Secondary Grade vs CGPA

Yes , the grades in college have an impact on a student's performance in university.

6.1.5 Working on the data of Lahore Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affects CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## PROG_CODE     6   11.3  1.8766   10.29 2.39e-11 ***
## Residuals  4619  842.4  0.1824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 102. ANOVA - Degree vs CGPA

With p value smaller than significance value , we can state that no students' performance does not vary for each degree program .

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affects CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## GENDER        1   24.2  24.150     133 <2e-16 ***
## Residuals  2398  435.6   0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 103. ANOVA - Gender vs CGPA

From the results of ANOVA , as p value is smaller than significance value we can clearly see CGPA does not relate to gender.

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affects CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## CITY          7   4.94  0.7061   4.144 0.000257 ***
## Residuals   232  39.53  0.1704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 104. ANOVA - City vs CGPA

Since p value and significance value are equal we go with the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Faisalabad	81	2.06	2.46	2.64	2.73	2.99	3.71
Gujranwala	60	2.06	2.41	2.64	2.68	2.97	3.50
Lahore	2472	2.00	2.50	2.80	2.84	3.15	3.98
Multan	62	2.00	2.45	2.61	2.67	2.95	3.44
Sahiwal	34	2.14	2.37	2.66	2.72	3.05	3.45
Sargodha	44	2.13	2.36	2.51	2.61	2.87	3.50
Sialkot	33	2.02	2.42	2.63	2.84	3.22	3.94

Table 19. Numerical Summary of Cities - Lahore Campus

Students from Lahore and Sialkot cities perform better in the Lahore Campus .

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary education background affects CGPA

Alternative Hypothesis 4: Secondary education background doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## SECONDARY     1    5.1   5.148   26.01 3.81e-07 ***
## Residuals  1598  316.3   0.198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 105. ANOVA - Secondary Education vs CGPA

With p value smaller than significance value we can say that there is no difference between performance of students based on whether they belong to a matric background or an O levels background.

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affects CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL       28   29.5  1.0525   5.821 <2e-16 ***
## Residuals  4349  786.4  0.1808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 106. ANOVA - School vs CGPA

No , depending on the school from which a student studied , his/her performance does not vary .

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary education background affects CGPA
Alternative Hypothesis 6: Higher Secondary education background doesn't affect CGPA .

```
##                  Df Sum Sq Mean Sq F value  Pr(>F)
## HIGHER_SECONDARY    1   2.02  2.0186   10.48 0.00123 **
## Residuals        1398 269.24  0.1926
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 107. ANOVA - Higher Secondary Education vs CGPA

With a p value greater than the significance value , it shows that there is a difference in performance depending on the higher secondary education background .

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	700	2.03	2.59	2.92	2.94	3.27	4
HSSC	3889	2.00	2.50	2.79	2.83	3.13	4

Table 20. Numerical Summary of Higher Secondary Education - Lahore Campus

We can see that students of A level, although being in the minority, have a better average than the students from HSC . So we can say that A level students are better performing than students from HSC .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affects CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## COLLEGE      18   14.4  0.7978   4.389 1.6e-09 ***
## Residuals  4452  809.3  0.1818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 108. ANOVA - College vs CGPA

Since p value is smaller than significance value , we reject the null hypothesis

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affects CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM    14   8.37  0.5976   3.465 1.35e-05 ***
## Residuals  1470 253.53  0.1725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 109. ANOVA - Admission Year vs CGPA

With a smaller p value , no, the admission semester does not make a difference in performance .

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduating Year affects CGPA
Alternative Hypothesis 9: Graduating year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM     28  54.57  1.9490    15.2 <2e-16 ***
## Residuals   841 107.81  0.1282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 110. ANOVA - Graduation Year vs CGPA

No , the semester of graduation has no impact on the student performance.

Question 10 : Are grades in school important for a better performance at university ?

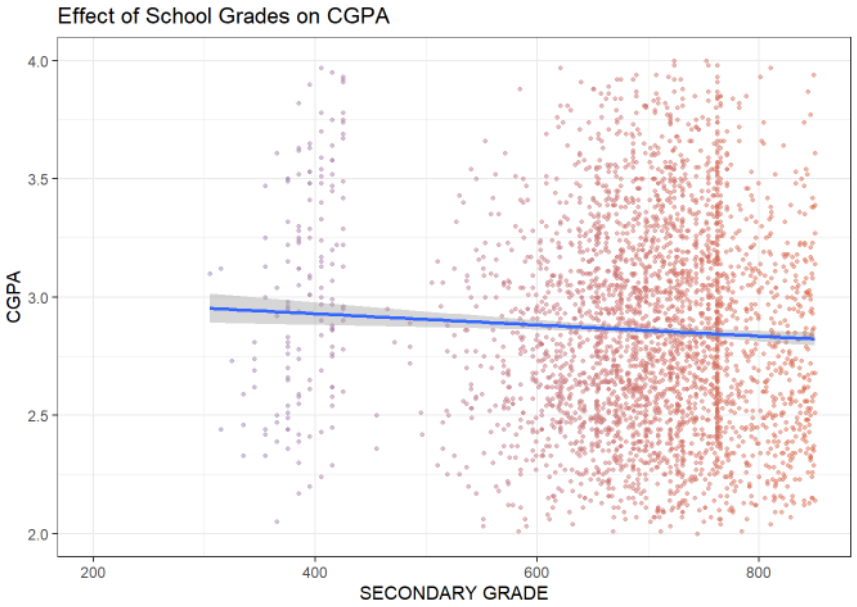


Figure 111. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  LAHORE_data$SEC_GRADE and LAHORE_data$CGPA
## t = -3.0256, df = 4624, p-value = 0.002495
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.07317419 -0.01565170
## sample estimates:
##         cor
## -0.04444979
```

Figure 112. Pearson Correlation - Secondary Grade vs CGPA

For Islamabad Campus , there is a negative impact of good school grades on CGPA at university .

Question 11 : Are grades in college important for a better performance at university ?

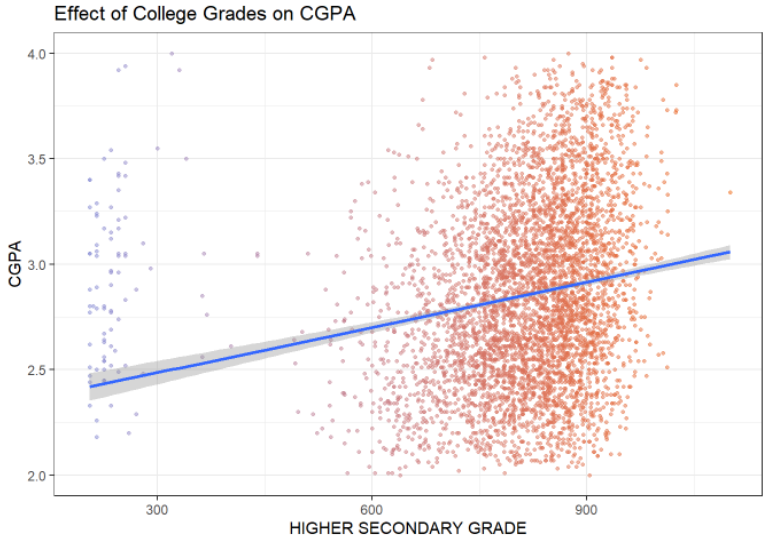


Figure 113. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  LAHORE_data$HIG_SEC_GRADE and LAHORE_data$CGPA
## t = 12.049, df = 4624, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1463891 0.2022723
## sample estimates:
##       cor
## 0.1744712
```

Figure 114. Pearson Correlation - Higher Secondary Grade vs CGPA

Yes , the grades in college have an impact on a student's performance in university.

6.1.6 Working on the data of Peshawar Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affect CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## PROG_CODE    3  1.726  0.5755   3.083 0.0291 *
## Residuals   156 29.122  0.1867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 115. ANOVA - Degree vs CGPA

Here the null hypothesis is accepted.

PROG_CODE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
BBA	74	2.00	2.62	3.00	3.01	3.37	3.93
BS(CS)	365	2.01	2.34	2.71	2.75	3.06	3.92
BS(EE)	487	2.07	2.50	2.87	2.88	3.20	3.99
BS(TE)	205	2.00	2.67	2.96	2.99	3.28	3.86

Table 21. Numerical Summary of Degree Programs - Peshawar Campus

Students from BBA outperform other students .

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affect CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## GENDER       1   2.85  2.8489   13.67 0.000283 ***
## Residuals   198  41.27  0.2085
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 116. ANOVA - Gender vs CGPA

Rejecting the null hypothesis.

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affect CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## CITY         2  0.918  0.4592    2.25  0.112
## Residuals   87 17.757  0.2041
```

Figure 117. ANOVA - City vs CGPA

As p value is greater than significance value we accept the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Abbottabad	11	2.00	2.51	2.73	2.79	3.00	3.77
Bahawalpur	11	2.16	2.59	2.77	2.90	3.25	3.56
Dera Ghazi Khan	16	2.16	2.47	2.72	2.75	3.04	3.26
Faisalabad	15	2.12	2.29	2.55	2.70	2.96	3.77
Gilgit	12	2.10	2.63	2.92	2.81	3.05	3.39
Islamabad	44	2.01	2.35	2.76	2.72	3.01	3.67
Kohat	19	2.14	2.44	2.88	2.84	3.16	3.66
Lahore	66	2.07	2.27	2.68	2.69	2.97	3.76
Layyah	12	2.15	2.78	3.00	3.07	3.39	3.75
Mardan	29	2.08	2.57	2.87	2.90	3.15	3.79
Multan	15	2.12	2.36	2.64	2.78	3.23	3.68
Nowshera	24	2.23	2.52	2.91	2.85	3.02	3.57
Others	21	2.14	2.57	2.66	2.72	2.91	3.59
Peshawar	524	2.00	2.54	2.88	2.91	3.22	3.99
Quetta	15	2.60	3.10	3.55	3.36	3.70	3.86
Rawalpindi	29	2.04	2.35	2.65	2.79	3.25	3.67
Sargodha	14	2.30	2.63	2.88	2.93	3.12	3.76
Swabi	22	2.24	2.59	2.91	2.87	3.05	3.61

Table 22. Numerical Summary of Cities - Peshawar Campus

Students from Layyah have a better average CGPA than students from other cities .

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary Education (SSC/O Level) affect CGPA
Alternative Hypothesis 4: Secondary Education (SSC/O Level) doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY    1  0.499  0.4987   2.412  0.126
## Residuals   58 11.990  0.2067
```

Figure 118. ANOVA - Secondary Education vs CGPA

Clearly , accept the null hypothesis.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	37	2.02	2.76	3.04	3.09	3.43	3.99
SSC	1079	2.00	2.49	2.84	2.86	3.19	3.95

Table 23. Numerical Summary of Secondary Education - Peshawar Campus

O Level students have a better performance than SSC students.

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affect CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL       15  4.696  0.3131   1.596 0.0817 .
## Residuals   144 28.257  0.1962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 119. ANOVA - School vs CGPA

The null hypothesis is acceptable.

saint marys academy	16	2.20	2.48	2.92	2.87	3.21	3.67
saint paul	12	2.28	2.40	2.54	2.61	2.78	3.23
st marys high school khyber road peshawar	20	2.16	2.60	2.75	2.81	3.02	3.41
st pauls english high school	31	2.10	2.90	3.07	3.06	3.36	3.79
university wensam college	78	2.00	2.67	2.99	2.98	3.26	3.99
usman science school	297	2.02	2.69	2.98	2.98	3.24	3.95

Table 24. Numerical Summary of Schools - Peshawar Campus

Students from St Paul's English High School have a better performance .

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary Education (HSSC / A Level) affect CGPA
Alternative Hypothesis 6: Higher Secondary Education (HSSC / A Level) doesn't affect CGPA .

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY  1  1.193  1.1929   6.163  0.016 *
## Residuals        58 11.227  0.1936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 120. ANOVA - Higher Secondary Education vs CGPA

The null hypothesis is accepted.

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	34	2.02	2.72	3.00	3.02	3.39	3.99
HSSC	1081	2.00	2.49	2.85	2.86	3.19	3.95

Table 25. Numerical Summary of Higher Secondary Education - Peshawar Campus

Students from A Level have a better performance .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affect CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      14  5.362  0.3830   2.013  0.021 *
## Residuals   135 25.688  0.1903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 121. ANOVA - College vs CGPA

The acceptance of the null hypothesis is evident.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affect CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM    13  11.51  0.8856   5.147 1.68e-08 ***
## Residuals   406  69.86  0.1721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 122. ANOVA - Admission Year vs CGPA

A clear rejection of the null hypothesis.

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduation Year affect CGPA
Alternative Hypothesis 9: Graduation Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## LAST_SEM     13  15.48  1.1905   7.075 1.94e-12 ***
## Residuals   406  68.32  0.1683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 123. ANOVA - Graduation Year vs CGPA

Rejected with p value being less than significance value.

Question 10 : Are grades in school important for a better performance at university ?

For the remaining attributes we calculate correlation coefficient .

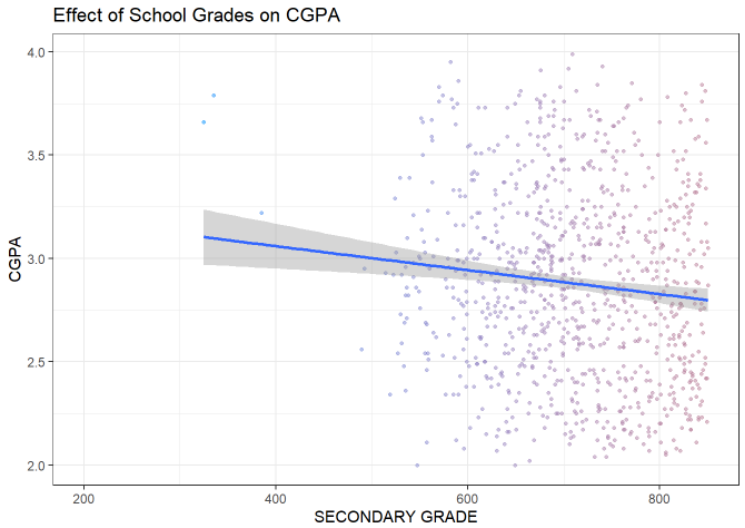


Figure 124. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  PESHAWAR_data$SEC_GRADE and PESHAWAR_data$CGPA
## t = -3.588, df = 1129, p-value = 0.0003476
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.16345823 -0.04818631
## sample estimates:
##       cor
## -0.106179
```

Figure 125. Pearson Correlation - Secondary Grade vs CGPA

Negative correlation between school grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

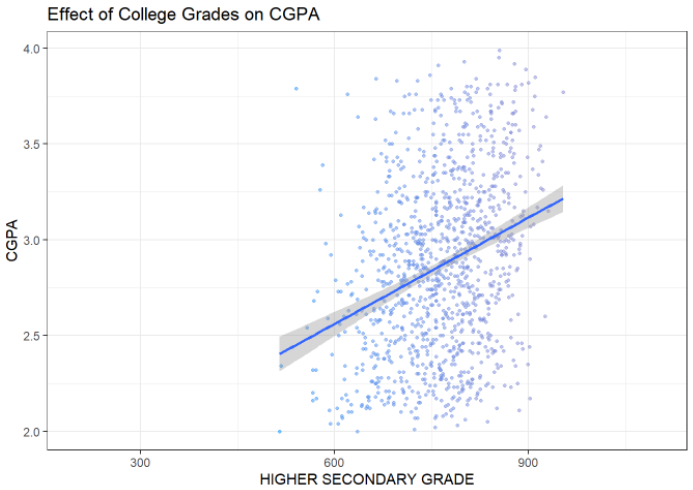


Figure 126. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  PESHAWAR_data$HIG_SEC_GRADE and PESHAWAR_data$CGPA
## t = 9.3458, df = 1129, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2130063 0.3212430
## sample estimates:
##     cor
## 0.26797
```

Figure 127. Pearson Correlation - Higher Secondary Grade vs CGPA

Not a strong relationship between college grade and CGPA.

6.2 Considering the data of warning students as a part of analysis

6.2.1 Working on the whole data

Question 1 : Do students from one campus perform better as compared to students from other campuses ? OR which campus has a higher average CGPA ?

Null Hypothesis 1: Campus affect CGPA
Alternative Hypothesis 1: Campus doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## CAMPUS        2   0.90  0.4479   5.911 0.00287 **
## Residuals   597  45.24  0.0758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 128. ANOVA - Campus vs CGPA

With p value greater than significance value it can be seen that the null hypothesis is clearly accepted .

CAMPUS	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ISB	707	1.96	2.19	2.35	2.40	2.56	3.59
KHI	1	1.96	1.96	1.96	1.96	1.96	1.96
LHR	846	2.00	2.24	2.40	2.44	2.60	3.80
PWR	226	2.00	2.28	2.49	2.49	2.66	3.61

Table 26. Numerical Summary of Campuses for whole data

Students from Peshawar campus perform better under warning .

Question 2 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 2: Degree Program affect CGPA
Alternative Hypothesis 2: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## PROG_CODE     3    3.5  1.1668   15.68 4.61e-10 ***
## Residuals  1722  128.1  0.0744
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 129. ANOVA - Degree vs CGPA

This null hypothesis is rejected with p value less than significance value.

Question 3 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 3: Gender affect CGPA
Alternative Hypothesis 3: Gender doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## GENDER        1   1.18   1.176   14.17 0.000189 ***
## Residuals   458  38.02   0.083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 130. ANOVA - Gender vs CGPA

And clearly CGPA doesn't relate to gender.

Question 4 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 4: City affect CGPA
Alternative Hypothesis 4: City doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## CITY        20  1.782 0.08908   0.992  0.481
## Residuals   84  7.546 0.08983
```

Figure 131. ANOVA - City vs CGPA

Comparing the p value and significance value , we accept the null hypothesis .

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Abbottabad	17	2.05	2.18	2.24	2.32	2.48	2.77
Attock	10	2.00	2.24	2.51	2.58	2.80	3.61
Faisalabad	15	2.02	2.17	2.31	2.40	2.64	2.89
Islamabad	308	2.00	2.19	2.34	2.40	2.58	3.30
Lahore	162	2.03	2.30	2.46	2.50	2.66	3.26
Multan	16	2.00	2.13	2.25	2.27	2.42	2.66
Others	719	2.00	2.23	2.37	2.42	2.56	3.80
Peshawar	120	2.04	2.30	2.55	2.52	2.70	3.31
Rawalpindi	193	1.96	2.20	2.37	2.40	2.55	3.36
Sargodha	14	2.01	2.17	2.28	2.29	2.43	2.58
Wah Cantt	13	2.06	2.23	2.49	2.47	2.66	2.89

Table 27. Numerical Summary of Cities for whole data

Students from Attock perform better under warning .

Question 5 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 5: Secondary Education (SSC/O Level) affect CGPA
Alternative Hypothesis 5: Secondary Education (SSC/O Level) doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY    1   0.01 0.01240   0.167  0.683
## Residuals  498  37.07 0.07443
```

Figure 132. ANOVA - Secondary Education vs CGPA

With p level almost equal to significance value we accept the null hypothesis.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	259	1.96	2.20	2.37	2.41	2.58	3.80
SSC	1425	1.96	2.22	2.39	2.43	2.60	3.72

Table 28. Numerical Summary of Secondary Education for whole data

Under warning students of SSC perform better .

Question 6 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 6: School affect CGPA
Alternative Hypothesis 6: School doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL      13  0.689 0.05300   0.623  0.831
## Residuals  126 10.720 0.08508
```

Figure 133. ANOVA - School vs CGPA

As can be seen that p value is greater than significance level , clearly accept the null hypothesis.

Question 7 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 7: Higher Secondary Education (HSSC / A Level) affect CGPA
Alternative Hypothesis 7: Higher Secondary Education (HSSC / A Level) doesn't affect CGPA .

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY  1   0.00 0.00308   0.037  0.847
## Residuals       398  32.82 0.08247
```

Figure 134. ANOVA - Higher Secondary Education vs CGPA

With a p value greater than the significance value , we accept the null hypothesis.

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	200	2.00	2.19	2.37	2.41	2.54	3.80
HSSC	1506	1.96	2.23	2.39	2.43	2.61	3.63

Table 29. Numerical Summary of Higher Secondary Education for whole data

For warning students , students with HSC have a better performance.

Question 8 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 8: College affect CGPA
Alternative Hypothesis 8: College doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE     10  0.968 0.09676   1.406  0.189
## Residuals   99  6.811 0.06880
```

Figure 135. ANOVA - College vs CGPA

Since p value is greater than significance value , the null hypothesis is accepted.

Question 9 : Does a semester of admission impact performance ?

Null Hypothesis 9: Admission Year affect CGPA
Alternative Hypothesis 9: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM     9   2.78 0.30858   4.077 3.39e-05 ***
## Residuals  1723 130.42 0.07569
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 136. ANOVA - Admission Year vs CGPA

With a very small p value , there is evidence against the null hypothesis hence rejected.

Question 10 : Does a graduating semester impact student performance ?

Null Hypothesis 10: Graduation Year affect CGPA
Alternative Hypothesis 10: Graduation Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## LAST_SEM     12   3.10 0.25842   3.376 7.08e-05 ***
## Residuals  1294  99.06 0.07656
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 137. ANOVA - Graduation Year vs CGPA

Since the p value is less than the significance value , we will go with the alternative hypothesis.

For the remaining attributes we calculate correlation coefficient .

Question 11 : Are grades in school important for a better performance at university ?

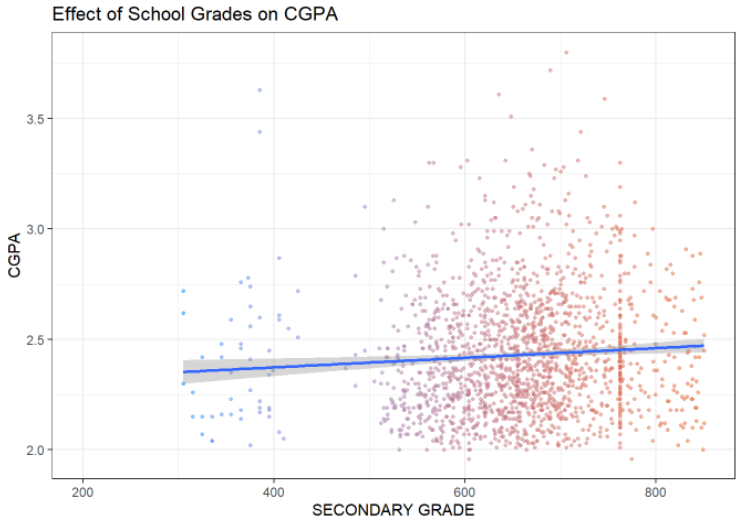


Figure 138. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  data$SEC_GRADE and data$CGPA
## t = 3.6131, df = 1778, p-value = 0.000311
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03906798 0.13131479
## sample estimates:
##        cor
## 0.08537433
```

Figure 139. Pearson Correlation - Secondary Grade vs CGPA

The value of 0.08 shows that there is no significant correlation between Secondary Grade and CGPA.

Question 12 : Are grades in college important for a better performance at university ?

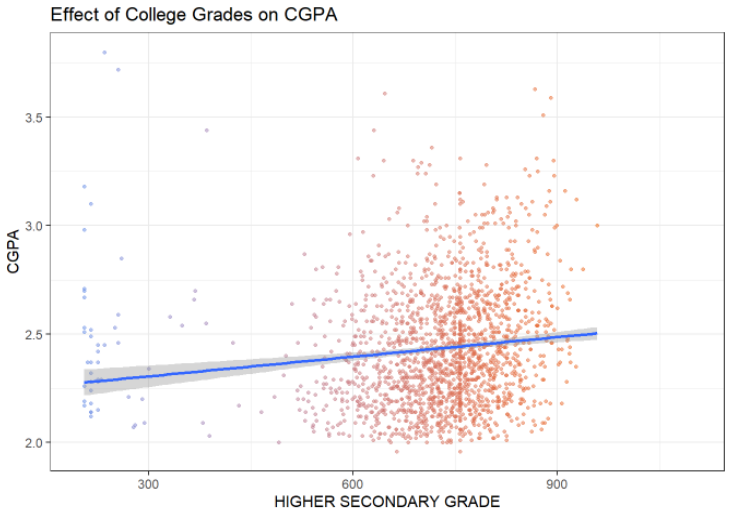


Figure 140. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  data$HIG_SEC_GRADE and data$CGPA
## t = 5.165, df = 1778, p-value = 2.674e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07554868 0.16710062
## sample estimates:
##       cor
## 0.1215832
```

Figure 141. Pearson Correlation - Higher Secondary Grade vs CGPA

Higher Secondary Grade and CGPA show a weak correlation.

6.2.2 Working on the data of Faisalabad Campus

Warning data for Faisalabad Campus wasn't provided, therefore cannot perform statistical analysis on Faisalabad's data.

6.2.3 Working on the data of Islamabad Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affect CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## PROG_CODE     3    3.5  1.1668   15.68 4.61e-10 ***
## Residuals  1722  128.1  0.0744
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 142. ANOVA - Degree vs CGPA

Here the null hypothesis is rejected.

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affect CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## GENDER       1   0.67  0.6705   9.773 0.00204 **
## Residuals  198  13.59  0.0686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 143. ANOVA - Gender vs CGPA

Accepting the null hypothesis.

GENDER	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
F	102	1.96	2.28	2.47	2.49	2.67	3.29
M	605	2.00	2.18	2.34	2.39	2.55	3.59

Table 30. Numerical Summary of Gender - Islamabad Campus

Female students of Islamabad perform better under warning .

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affect CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## CITY         4  0.316 0.07908   1.042  0.396
## Residuals   45  3.414 0.07587
```

Figure 144. ANOVA - City vs CGPA

As p value is greater than significance value we accept the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Abbottabad	15	2.05	2.19	2.24	2.32	2.44	2.77
Islamabad	299	2.00	2.19	2.34	2.40	2.58	3.30
Others	42	2.02	2.17	2.30	2.34	2.43	3.06
Rawalpindi	189	1.96	2.19	2.37	2.40	2.55	3.36
Wah Cantt	13	2.06	2.23	2.49	2.47	2.66	2.89

Table 31. Numerical Summary of Cities - Islamabad Campus

Students from Wah Cantt have a better performance than students from other cities .

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary Education (SSC/O Level) affect CGPA
Alternative Hypothesis 4: Secondary Education (SSC/O Level) doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY     1  0.062 0.06175   0.785  0.377
## Residuals   168 13.219 0.07869
```

Figure 145. ANOVA - Secondary Education vs CGPA

Clearly , accept the null hypothesis.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	89	2.00	2.18	2.36	2.39	2.55	3.08
SSC	602	1.96	2.19	2.35	2.40	2.57	3.59

Table 32. Numerical Summary of Secondary Education - Islamabad Campus

Even in Islamabad Campus , under warning SSC students perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affect CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL      12  0.942 0.07853   0.953  0.504
## Residuals   52  4.285 0.08240
```

Figure 146. ANOVA - School vs CGPA

The null hypothesis is acceptable.

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa

Null Hypothesis 6: Higher Secondary Education (HSSC / A Level) affect CGPA
Alternative Hypothesis 6: Higher Secondary Education (HSSC / A Level) doesn't affect CGPA .

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY   1  0.025 0.02483   0.376   0.54
## Residuals        148  9.763 0.06597
```

Figure 147. ANOVA - Higher Secondary Education vs CGPA

The null hypothesis is accepted.

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	75	2.01	2.18	2.34	2.37	2.54	3.08
HSSC	587	1.96	2.19	2.35	2.40	2.57	3.59

Table 33. Numerical Summary of Higher Secondary Education - Islamabad Campus

When in warning, the performance of HSC students is better .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affect CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      8  0.805 0.10062   1.716  0.107
## Residuals   81  4.749 0.05863
```

Figure 148. ANOVA - College vs CGPA

The acceptance of the null hypothesis is evident.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affect CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## FIRST_SEM    13  11.51  0.8856   5.147 1.68e-08 ***
## Residuals   406  69.86  0.1721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 149. ANOVA - Admission Year vs CGPA

A clear rejection of the null hypothesis.

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduation Year affect CGPA
Alternative Hypothesis 9: Graduation Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## LAST_SEM     19   3.06 0.16113   2.204 0.00234 **
## Residuals   638  46.64 0.07311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 150. ANOVA - Graduation Year vs CGPA

Accepted with p value being greater than significance value.

Spring 2006	16	2.08	2.22	2.40	2.45	2.66	2.92
Spring 2007	16	2.03	2.24	2.27	2.35	2.42	2.83
Spring 2008	43	2.06	2.31	2.45	2.47	2.68	3.12
Spring 2009	79	2.06	2.21	2.38	2.43	2.57	3.36
Spring 2010	38	2.01	2.17	2.36	2.44	2.66	3.29
Spring 2011	54	2.00	2.19	2.34	2.38	2.52	3.06
Spring 2012	42	2.01	2.23	2.45	2.51	2.75	3.59
Spring 2013	50	2.12	2.19	2.37	2.40	2.49	3.12
Spring 2014	50	2.00	2.23	2.41	2.45	2.61	3.23
Summer 2010	14	2.01	2.19	2.30	2.29	2.38	2.57
Summer 2011	23	2.04	2.13	2.25	2.29	2.43	2.59
Summer 2012	22	2.01	2.17	2.25	2.34	2.51	2.97
Summer 2014	12	2.00	2.04	2.27	2.26	2.38	2.67

Table 34. Numerical Summary of Semesters - Islamabad Campus

Performance of Spring 2012 was better .

Question 10 : Are grades in school important for a better performance at university ?

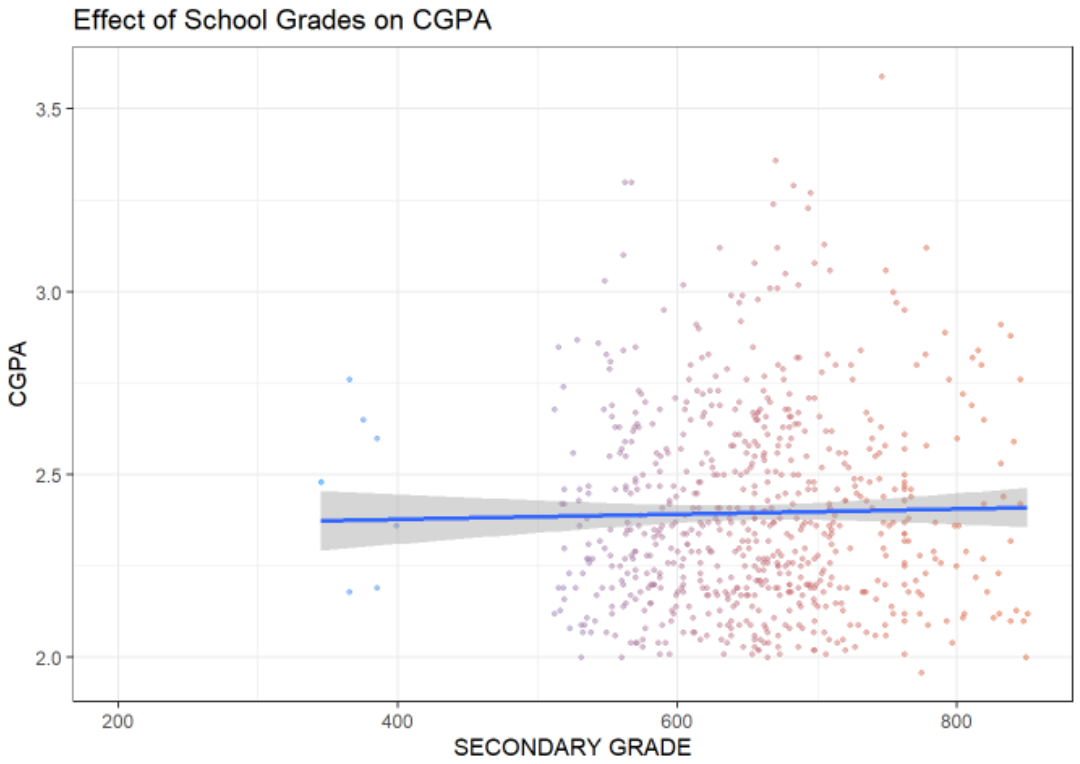


Figure 151. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  ISLAMABAD_data$SEC_GRADE and ISLAMABAD_data$CGPA
## t = 1.4516, df = 705, p-value = 0.1471
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.01922376  0.12780889
## sample estimates:
##       cor
## 0.05458848
```

Figure 152. Pearson Correlation - Secondary Grade vs CGPA

Very weak correlation between school grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

```
##
##  Pearson's product-moment correlation
##
## data:  ISLAMABAD_data$HIG_SEC_GRADE and ISLAMABAD_data$CGPA
## t = 3.0527, df = 705, p-value = 0.002353
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.04082729 0.18638346
## sample estimates:
##       cor
## 0.1142183
```

Figure 153. Pearson Correlation - Higher Secondary Grade vs CGPA

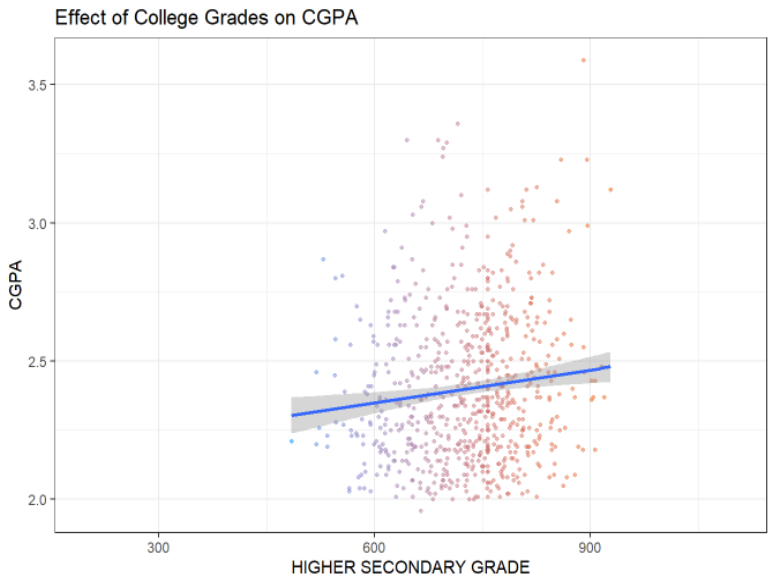


Figure 154. Correlation Graph - Higher Secondary Grade vs CGPA

Not a strong relationship between college grade and CGPA.

6.2.4 Working on the data of Karachi Campus

This analysis can't be performed for warning students data of Karachi campus , since there was only a single warning student in Karachi data .

6.2.5 Working on the data of Lahore Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affect CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## PROG_CODE     3  1.312  0.4373   6.304 0.00042 ***
## Residuals   196 13.598  0.0694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 155. ANOVA - Degree vs CGPA

Here the null hypothesis is rejected.

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affect CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## GENDER        1  0.622  0.6222   7.657 0.0061 **
## Residuals   238 19.341  0.0813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 156. ANOVA - Gender vs CGPA

Accepting the null hypothesis.

GENDER	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
F	122	2.03	2.34	2.47	2.53	2.66	3.72
M	724	2.00	2.23	2.38	2.43	2.58	3.80

Table 35. Numerical Summary of Gender - Lahore Campus

Female students with warning perform better .

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affect CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## CITY          2   1.20  0.5986   8.208 0.000296 ***
## Residuals   801  58.42  0.0729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 157. ANOVA - City vs CGPA

As p value is smaller than significance value we reject the null hypothesis.

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary Education (SSC/O Level) affect CGPA
Alternative Hypothesis 4: Secondary Education (SSC/O Level) doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY     1  0.056 0.05592   0.661  0.417
## Residuals   318 26.892 0.08457
```

Figure 158. ANOVA - Secondary Education vs CGPA

Clearly , accept the null hypothesis.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	89	2.00	2.18	2.36	2.39	2.55	3.08
SSC	602	1.96	2.19	2.35	2.40	2.57	3.59

Table 36. Numerical Summary of Secondary Education - Lahore Campus

Even in Lahore Campus , under warning SSC students perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affect CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL       3 0.0256 0.00854   0.122  0.947
## Residuals   36 2.5242 0.07012
```

Figure 159. ANOVA - School vs CGPA

The null hypothesis is acceptable.

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary Education (HSSC / A Level) affect CGPA
Alternative Hypothesis 6: Higher Secondary Education (HSSC / A Level) doesn't affect CGPA .

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## HIGHER_SECONDARY   1  0.013 0.01305   0.138  0.711
## Residuals        238 22.546 0.09473
```

Figure 160. ANOVA - Higher Secondary Education vs CGPA

The null hypothesis is accepted.

HIGHER_SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
ALEVEL	125	2.00	2.23	2.37	2.44	2.54	3.80
HSSC	699	2.02	2.25	2.40	2.45	2.61	3.63

Table 37. Numerical Summary of Higher Secondary Education - Lahore Campus

HSC students perform better under warning .

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affect CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## COLLEGE      3  0.721  0.2405   2.304 0.0934 .
## Residuals   36  3.758  0.1044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 161. ANOVA - College vs CGPA

The acceptance of the null hypothesis is evident.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affect CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## FIRST_SEM     9  1.374  0.1526   1.882 0.0544 .
## Residuals   290 23.518  0.0811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 162. ANOVA - Admission Year vs CGPA

A clear acceptance of the null hypothesis.

FIRST_SEM	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Fall 2001	46	2.03	2.22	2.33	2.41	2.51	3.51
Fall 2002	71	2.04	2.25	2.41	2.45	2.61	3.80
Fall 2003	59	2.05	2.26	2.37	2.42	2.52	3.05
Fall 2004	89	2.09	2.27	2.36	2.43	2.52	3.72
Fall 2005	83	2.03	2.27	2.38	2.43	2.51	3.30
Fall 2006	110	2.07	2.25	2.43	2.47	2.61	3.63
Fall 2007	103	2.00	2.23	2.35	2.40	2.55	3.11
Fall 2008	104	2.02	2.18	2.33	2.38	2.55	3.16
Fall 2009	103	2.03	2.30	2.46	2.52	2.66	3.28
Fall 2010	55	2.13	2.41	2.51	2.58	2.76	3.24

Table 38. Numerical Summary of Semesters - Lahore Campus

Under warning , the performance in the Fall 2010 semester was better than other semesters.

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduation Year affect CGPA
Alternative Hypothesis 9: Graduation Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM     19  1.279 0.06730   0.972  0.494
## Residuals   280 19.385 0.06923
```

Figure 163. ANOVA - Graduation Year vs CGPA

Accepted with p value being greater than significance value.

Spring 2006	19	2.10	2.37	2.60	2.59	2.77	3.51
Spring 2007	17	2.03	2.16	2.24	2.35	2.46	3.05
Spring 2008	26	2.18	2.34	2.46	2.52	2.60	3.44
Spring 2009	48	2.05	2.28	2.37	2.43	2.51	3.72
Spring 2010	52	2.03	2.27	2.42	2.51	2.67	3.44
Spring 2011	48	2.08	2.24	2.46	2.44	2.56	3.11
Spring 2012	63	2.04	2.21	2.40	2.44	2.61	3.16
Spring 2013	86	2.03	2.24	2.49	2.50	2.66	3.28
Spring 2014	65	2.04	2.32	2.46	2.52	2.69	3.24

Table 39. Numerical Summary of Graduating Semesters - Lahore Campus

As for warning data , performance in Spring 2006 was better compared to the rest of the semesters.

Question 10 : Are grades in school important for a better performance at university ?

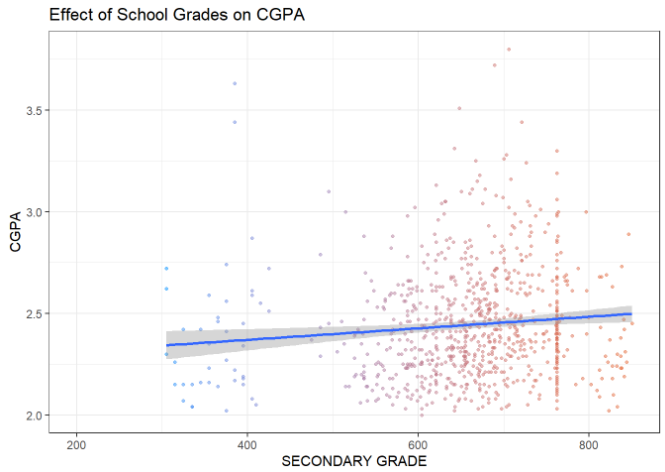


Figure 164. Correlation Graph - Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  LAHORE_data$SEC_GRADE and LAHORE_data$CGPA
## t = 3.0368, df = 844, p-value = 0.002464
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03682159 0.17017583
## sample estimates:
##      cor
## 0.103966
```

Figure 165. Pearson Correlation - Secondary Grade vs CGPA

Weak correlation between school grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

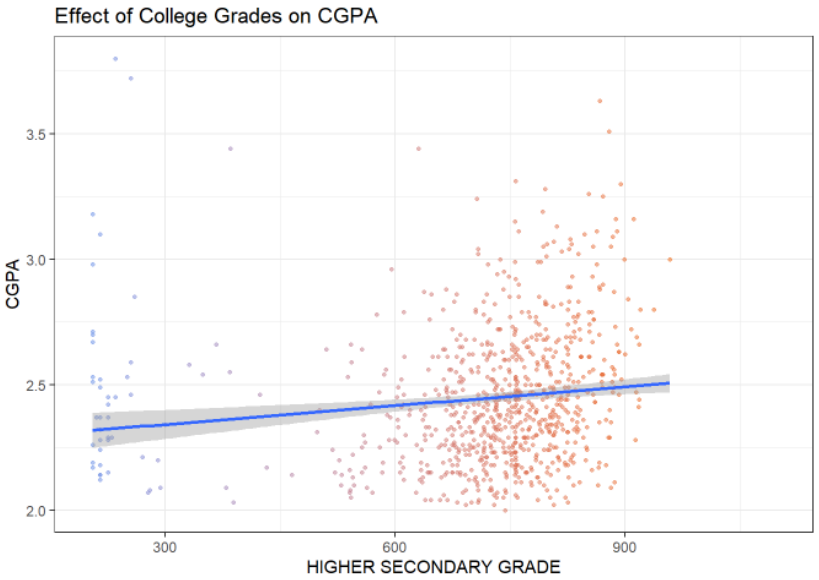


Figure 166. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  LAHORE_data$HIG_SEC_GRADE and LAHORE_data$CGPA
## t = 3.9159, df = 844, p-value = 9.732e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.06678281 0.19919279
## sample estimates:
##       cor
## 0.1335839
```

Figure 167. Pearson Correlation - Higher Secondary Grade vs CGPA

Not a strong relationship between college grade and CGPA.

6.2.6 Working on the data of Peshawar Campus

Question 1 : Is there any effect of the particular degree program on the performance of students ?

Null Hypothesis 1: Degree Program affect CGPA
Alternative Hypothesis 1: Degree Program doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## PROG_CODE     3  0.433  0.1442   1.754  0.157
## Residuals   222 18.248  0.0822
```

Figure 168. ANOVA - Degree vs CGPA

Here the null hypothesis is accepted.

PROG_CODE	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
BBA	14	2.20	2.32	2.66	2.60	2.77	3.15
BS(CS)	62	2.03	2.25	2.49	2.47	2.64	3.23
BS(EE)	22	2.12	2.41	2.55	2.59	2.70	3.10
BS(TE)	128	2.00	2.25	2.46	2.47	2.64	3.61

Table 40. Numerical Summary of Degree Programs - Peshawar Campus

Students of BBA perform better under warning.

Question 2 : What role does gender play in academic performance ? Do girls tend to perform better than boys or vice versa ?

Null Hypothesis 2: Gender affect CGPA
Alternative Hypothesis 2: Gender doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## GENDER        1  0.129 0.12936   1.562  0.213
## Residuals   224 18.551 0.08282
```

Figure 169. ANOVA - Gender vs CGPA

Accepting the null hypothesis.

GENDER	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
F	12	2.3	2.44	2.53	2.59	2.62	3.31
M	214	2.0	2.26	2.48	2.49	2.66	3.61

Table 41. Numerical Summary of Gender - Peshawar Campus

The performance of female students under warning is better than male students.

Question 3 : Are students from a particular city/district performing better ? Which city/district has a higher average performance ?

Null Hypothesis 3: City affect CGPA
Alternative Hypothesis 3: City doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## CITY         5  1.338 0.26762    3.65 0.00375 **
## Residuals   157 11.510 0.07331
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 170. ANOVA - City vs CGPA

As p value is greater than significance value we accept the null hypothesis.

CITY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Islamabad	9	2.00	2.20	2.35	2.34	2.48	2.74
Kohat	7	2.32	2.45	2.58	2.61	2.79	2.88
Lahore	24	2.03	2.19	2.36	2.41	2.62	3.23
Multan	6	2.00	2.07	2.15	2.14	2.21	2.28
Others	8	2.17	2.36	2.58	2.59	2.80	3.15
Peshawar	109	2.04	2.30	2.55	2.52	2.69	3.31

Table 42. Numerical Summary of Cities - Peshawar Campus

Students of Kohat relatively have a better performance.

Question 4 : Do students from a SSC background perform better than students with an O Levels background or vice versa ?

Null Hypothesis 4: Secondary Education (SSC/O Level) affect CGPA
Alternative Hypothesis 4: Secondary Education (SSC/O Level) doesn't affect CGPA .

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## SECONDARY    1  0.051 0.05090   0.603  0.438
## Residuals   217 18.331 0.08448
```

Figure 171. ANOVA - Secondary Education vs CGPA

Clearly , accept the null hypothesis.

SECONDARY	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
OLEVEL	3	2.46	2.58	2.69	2.62	2.70	2.71
SSC	216	2.00	2.26	2.49	2.49	2.66	3.61

Table 43. Numerical Summary of Secondary Education - Peshawar Campus

In Peshawar , even under warning O Levels students perform better .

Question 5 : Which school has a better performance than other schools ? Does a student's school make a difference in his/her performance at FAST ?

Null Hypothesis 5: School affect CGPA
Alternative Hypothesis 5: School doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SCHOOL        4  0.231 0.05765   0.682  0.605
## Residuals   194 16.396 0.08452
```

Figure 172. ANOVA - School vs CGPA

The null hypothesis is acceptable.

Question 6 : Do students from a HSC background have a better CGPA than students with an O Levels background or vice versa ?

Null Hypothesis 6: Higher Secondary Education (HSSC / A Level) affect CGPA
Alternative Hypothesis 6: Higher Secondary Education (HSSC / A Level) doesn't affect CGPA .

Cannot apply ANOVA test to analyze the effect of Higher Secondary on CGPA for warning data of Peshawar Campus because there are only students from HSSC.

Question 7 : Students from which college have a better CGPA , as compared to other colleges?

Null Hypothesis 7: College affect CGPA
Alternative Hypothesis 7: College doesn't affect CGPA .

Accepting the null hypothesis.

Question 8 : Does a semester of admission impact performance ?

Null Hypothesis 8: Admission Year affect CGPA
Alternative Hypothesis 8: Admission Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## FIRST_SEM     8  1.582 0.19780   2.469 0.0142 *
## Residuals   203 16.263 0.08011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 173. ANOVA - Admission Year vs CGPA

A clear acceptance of the null hypothesis.

FIRST_SEM	Student_Count	Min	First_Quartile	Median	Mean	Third_Quartile	Max
Fall 2001	16	2.16	2.35	2.50	2.54	2.62	3.23
Fall 2002	16	2.03	2.48	2.63	2.61	2.80	3.13
Fall 2003	27	2.06	2.31	2.49	2.53	2.70	3.31
Fall 2004	13	2.03	2.30	2.36	2.40	2.58	2.83
Fall 2005	40	2.00	2.14	2.29	2.37	2.59	3.28
Fall 2006	26	2.12	2.29	2.49	2.51	2.66	3.61
Fall 2007	40	2.07	2.21	2.41	2.46	2.59	3.15
Fall 2008	15	2.24	2.56	2.69	2.67	2.76	3.01
Fall 2009	19	2.09	2.39	2.53	2.53	2.66	3.10

Table 44. Numerical Summary of Semesters - Peshawar Campus

Performance in the Fall 2008 semester was better .

Question 9 : Does a graduating semester impact student performance ?

Null Hypothesis 9: Graduation Year affect CGPA
Alternative Hypothesis 9: Graduation Year doesn't affect CGPA .

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## LAST_SEM     12  2.108 0.17565    2.22  0.013 *
## Residuals   163 12.895 0.07911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 174. ANOVA - Graduation Year vs CGPA

Accepted with p value being less than significance value.

Question 10 : Are grades in school important for a better performance at university ?

```
##
##  Pearson's product-moment correlation
##
## data:  PESHAWAR_data$SEC_GRADE and PESHAWAR_data$CGPA
## t = 1.569, df = 224, p-value = 0.118
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02659763  0.23161363
## sample estimates:
##       cor
## 0.1042647
```

Figure 175. Pearson Correlation - Secondary Grade vs CGPA

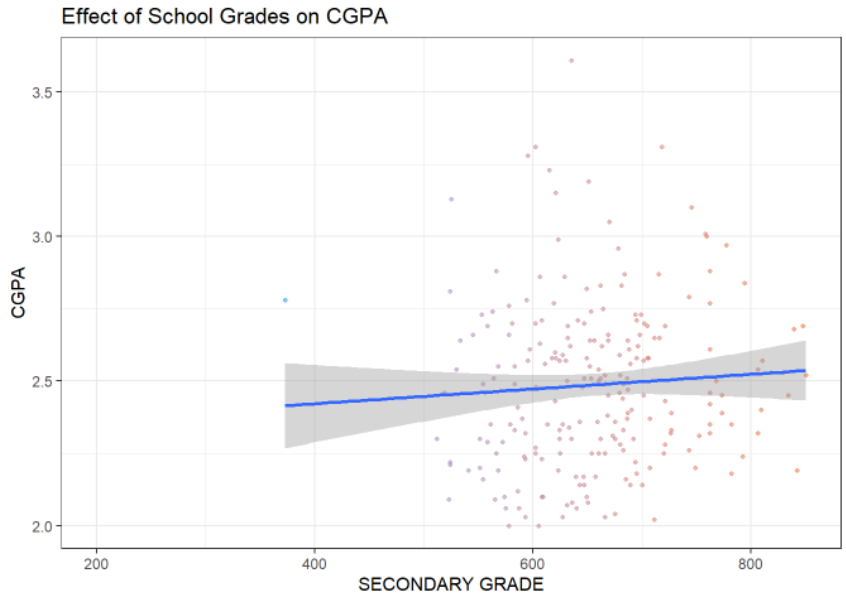


Figure 176. Correlation Graph - Secondary Grade vs CGPA

A weak correlation between school grade and CGPA.

Question 11 : Are grades in college important for a better performance at university ?

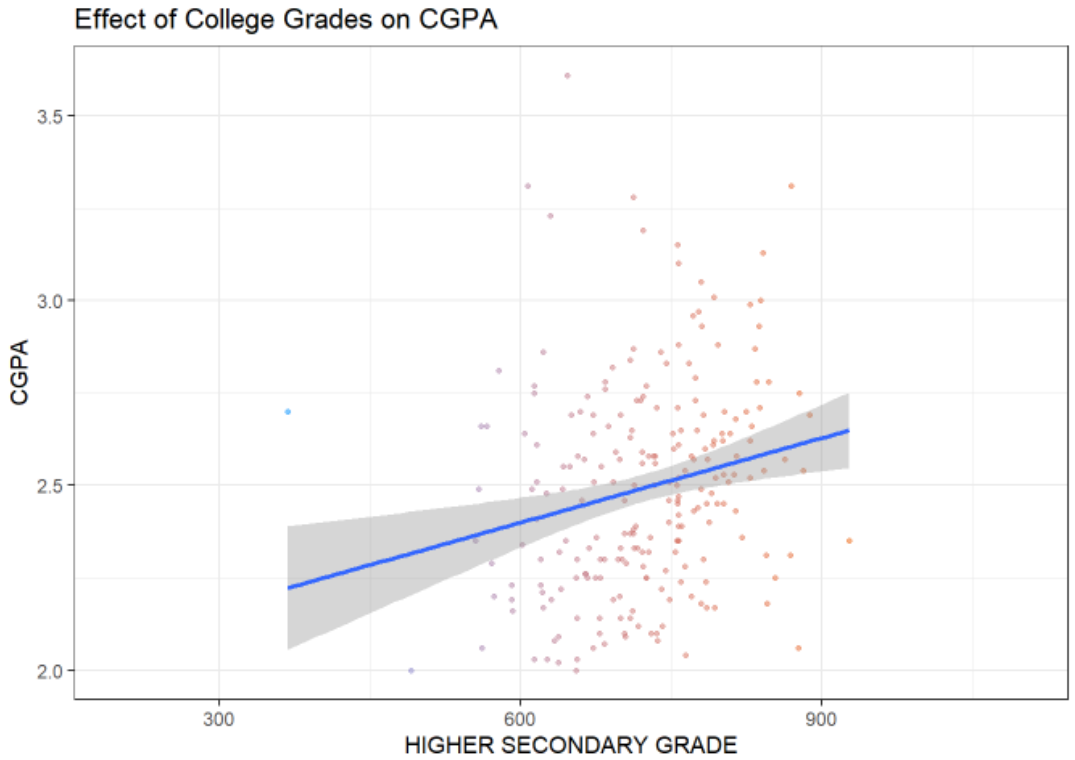


Figure 177. Correlation Graph - Higher Secondary Grade vs CGPA

```
##
##  Pearson's product-moment correlation
##
## data:  PESHAWAR_data$HIG_SEC_GRADE and PESHAWAR_data$CGPA
## t = 3.2676, df = 224, p-value = 0.001255
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.08516897 0.33448839
## sample estimates:
##       cor
## 0.2132986
```

Figure 178. Pearson Correlation - Higher Secondary Grade vs CGPA

Positive relationship between college grade and CGPA.

Chapter 7

RESULTS AND DISCUSSIONS

7.1 Statistical Analysis

From the above statistical analysis we summarize the correlations between CGPA and different attributes in the tables below :

7.1.1 Considering the data of warning students as an outlier (not a part of analysis)

Features	Evaluation Metric					
	Overall	**Faisalabad**	**Islamabad**	**Karachi**	**Lahore**	**Peshawar**
Degree	No	Yes	No	Yes	No	Yes
Gender	No	Yes	No	No	No	No
City	No	Yes	Yes	Yes	Yes	Yes
Secondary Education Background	Yes	Yes	No	Yes	No	Yes
School	Yes	Yes	Yes	Yes	Yes	Yes
Higher Secondary Education Background	Yes	Yes	Yes	Yes	No	Yes
College	Yes	Yes	Yes	Yes	Yes	Yes
Admission Year	Yes	Yes	No	No	Yes	No
Graduation Year	No	No	No	No	No	No
School Grade	No	No	Yes (neg)	Yes	Yes (neg)	Yes (neg)
College Grade	Yes	No	Yes	Yes	Yes	No

Table 45. Correlation Results without Warning Data

Table 45 shows how different attributes / features are correlated with CGPA for the data of students without warning .

7.1.2 Considering the data of warning students as a part of analysis

Features	Evaluation Metric					
	Overall	**Faisalabad**	**Islamabad**	**Karachi**	**Lahore**	**Peshawar**
Degree	No	-	No	-	No	Yes
Gender	No	-	Yes	-	Yes	Yes
City	Yes	-	Yes	-	No	Yes
Secondary Education Background	Yes	-	Yes	-	Yes	Yes
School	Yes	-	Yes	-	Yes	Yes
Higher Secondary Education Background	Yes	-	Yes	-	Yes	-
College	Yes	-	Yes	-	Yes	Yes
Admission Year	No	-	Yes	-	Yes	Yes
Graduation Year	No	-	Yes	-	Yes	Yes
School Grade	No	-	No	-	Yes	Yes
College Grade	Yes	-	Yes	-	Yes	Yes

Table 46. Correlation Results with Warning Data

Table 46 shows how different attributes / features are correlated with CGPA for the data of students with warning .

7.1.3 Important courses for each degree

Degree		
Computer Science	**Electrical Engineering**	**Business Administration**
Probability and Statistics	Digital Logical Design	Business Statistics
Database Systems	Interfacing and Programming	Fundamentals of Management
Introduction to Computing	Probability and Random Processes	Business Communication
Theory of Automata	Electromagnetic Theory	Financial Accounting
Operating Systems	Electro Mechanical Systems	Fundamentals of Accounting
Linear Algebra	Signals and Systems	

Table 47 Important Courses for Different Degree Programs

Table 47 shows the courses ,for the three degree programs CS , EE and BBA , which impact the overall performance of a student . A student performing better in these subjects i-e getting good grades , has a good CGPA .

7.2 Model Building

Since our project also focused on comparison between different regression and classification models , we implemented several models (the details of which are mentioned above in Chapter 4.5) . We got different results from those models , and the results are shown in the tables below :

7.2.1 Regression Models

Algorithm	Performance Metric	
	R squared	RMSE
Linear Regression	0.95882	0.08991
Polynomial Regression	0.96579	0.08377

Table 48. Performance of Regression Models

Table 48 shows performance of the regression models implemented , with Polynomial Regression showing a less RMSE .

7.2.2 Classification Models

Algorithm	Performance Metric
	Accuracy
Decision Tree	0.863
Gaussian Naive Bayes	0.791
K Nearest Neighbors	0.711
Logistic Regression	0.827
Random Forest	0.861
Support Vector Machine	0.847
Ensemble Classifier	0.847

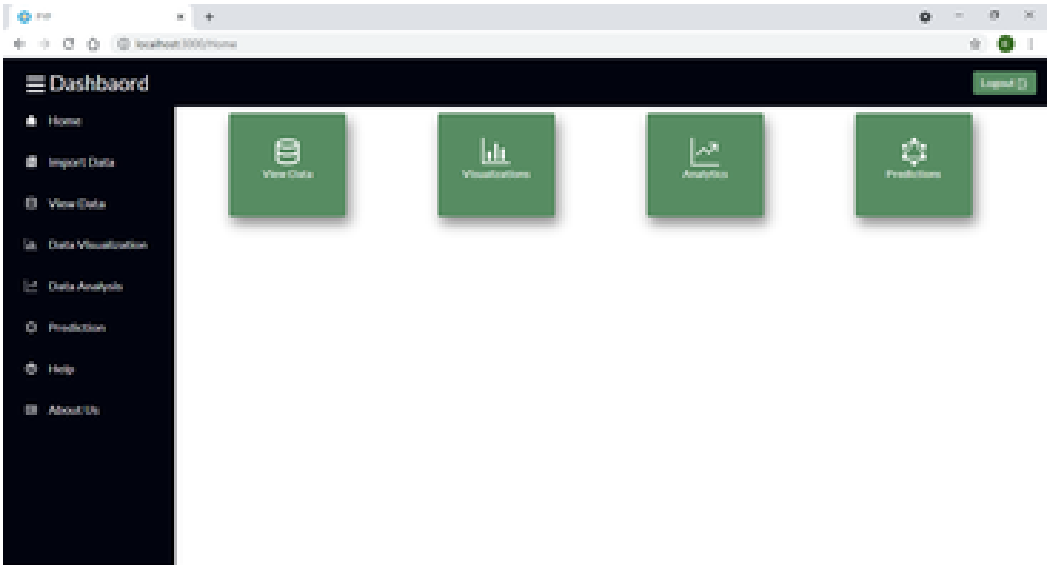
Table 49. Performance of Classification Models

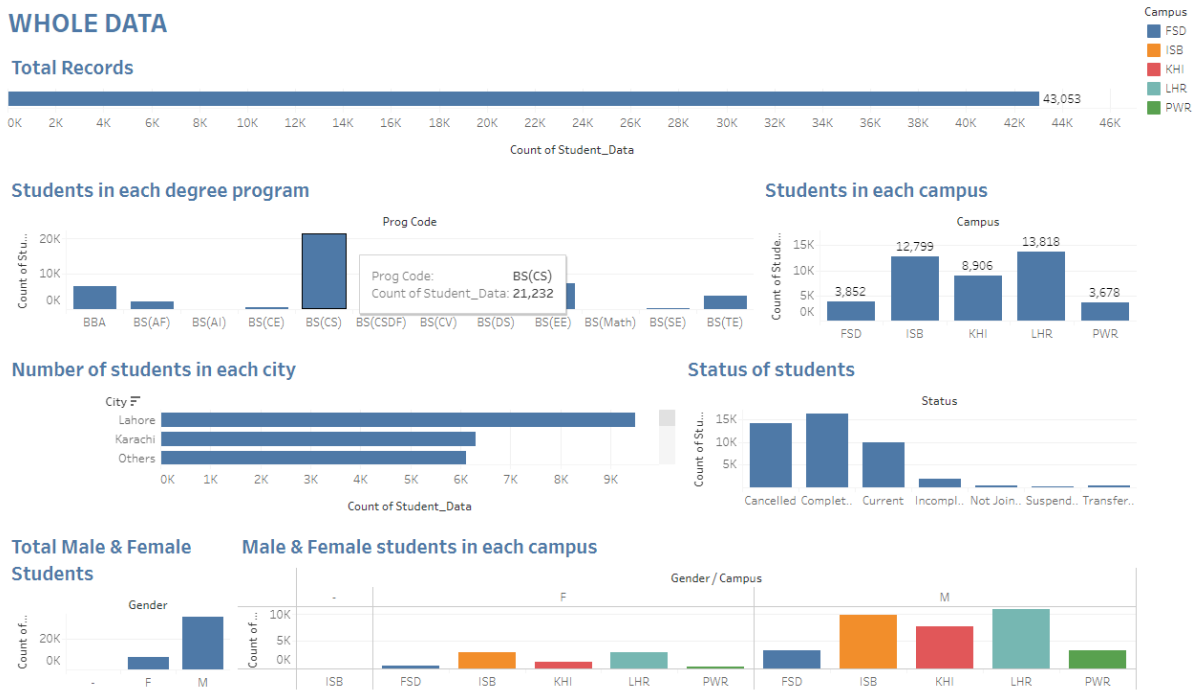
Table 49 shows accuracies of all the classification models implemented , with Decision Tree showing the highest accuracy . As for Ensemble Classifier , it was an ensemble of all the six algorithms mentioned above it in the table .

CHAPTER 8

DASHBOARD DEVELOPMENT

A dashboard was developed for visualization purposes. The dashboard shows the important visualization related to Exploratory Data Analysis and Statistical Analysis .





CHAPTER 9

CONCLUSION AND FUTURE WORK

A student's performance during his/her graduation cycle is highly affected by numerous factors . The factors such as student gender , the city to which a student belongs , the school from which a student studied , the college from which a student studied and the grades which the student obtained in his/her school and college greatly influence a student's performance. Also the educational background from which the student belongs i-e SSC or O Levels and HSC or A Levels is also a contributing factor to a student's performance.

Apart from these past attributes the student's performance is also related to attributes pertaining to his/her time at university . The degree program in which the student is enrolled , the courses in which a student is enrolled , a student's performance in important degree courses , a student's performance in each semester and attributes like warning count etc are also an indicator of how a student is performing . These factors are greatly related to the final CGPA of a student .

Apart from the attributes that we found through our analysis , there are several other attributes important for a student's performance. Further study can be made out to find out more of such attributes . Since our dataset contained only a certain number of attributes, we had to limit our analysis to the attributes discussed throughout the report . However the methodology and the process can be taken further to find out more about the problem and answer many more questions .

The website we have developed serves as a dashboard for the important visualizations. The working of the website can be continued further to develop a fully functional predictive analytics system that can be used by educational institutions for the purpose of importing their data and performing important steps in the data science process such as pre-processing , exploration , model building and so on .

REFERENCES

1. Hajra Waheed , Saeed-Ul Hassan , Julie Hardman , Salem Alelyani , Raheel Nawaz , ‘“Predicting academic performance of students from VLE big data using deep learning models” , Volume 104, March 2020
2. Eyman Alyahyan , Dilek , “Predicting academic success in higher education: literature review and best practices” , International Journal of Educational Technology in Higher Education , 10 February 2020
3. Muhammad Yunus Iqbal Basheer, Sofianita Mutalib, Nurzeatul Hamimah Abdul Hamid, Shuzlina Abdul-Rahman, Ariff Md Ab Malik , “Predictive analytics of university student intake using supervised methods” , IAES International Journal of Artificial Intelligence , Vol. 8, No. 4, December 2019, pp. 367~374
4. Sonali Rawat , “Predictive Analytics for Placement of Student- A Comparative Study” , International Research Journal of Engineering and Technology (IRJET) , Volume: 06 Issue: 06 , June 2019
5. Mukesh Kumar , Prof A.J.Singh , Dr Disha Handa , “Literature Survey on Student’s Performance Prediction in Education using Data Mining Techniques” , International Journal of Education and Management Engineering , November 2017
6. Ashish Dutt , Maizatul Akmar Ismail , Tutut Herawan , “A Systematic Review on Educational Data Mining” , IEEE Access , Volume: 5 , January 2017 .