



Enero 29, 2024

## Proyecto de Estadística

**Juan Antonio González**

juangonz@espol.edu.ec

ESPOL (FIEC)

**Ramón Ignacio Macías**

**Ochoa**

rimacias@espol.edu.ec

ESPOL (FIEC)

**Gary Steven Flores**

gsflores@espol.edu.ec

ESPOL (FIMCM)

**Germán David Correa**

gdcorrea@espol.edu.ec

ESPOL (FIMCM)

**Gabriel Cañarte Lucio**

lcanarte@espol.edu.ec

ESPOL (FIEC)

# Índice

1	Introducción .....	4
2	Objetivos .....	5
2.1	Objetivo General .....	5
2.2	Objetivos Específicos .....	5
3	Metodología .....	6
4	Variables de la base de Datos .....	7
4.1	Variables Cuantitativas .....	7
4.1.1	Estudiantes Blancos, Negros y Asiáticos .....	7
4.1.2	Promedios de SAT .....	8
4.1.3	Estudiantes Testados .....	9
4.2	Variables Cualitativas .....	10
4.2.1	Borough .....	10
4.2.2	City .....	12
5	Análisis y Resultados .....	13
5.1	Análisis de Correlación entre el promedio SAT y etnias .....	13
5.2	ANOVA ONE WAY entre variables promedio SAT y City .....	15
6	Conclusiones .....	16
6.1	Correlación entre la diversidad étnica y las puntuaciones SAT .....	16
6.2	Análisis de varianza entre promedios SAT y ubicación .....	17
7	Anexos .....	18
7.1	Código Fuente .....	18
	Bibliografía .....	19

## Índice de Figuras

Figura 1: Diagrama de cajas de los porcentajes de estudiantes blancos, negros y asiáticos. ....	7
Figura 2: Diagrama de cajas de los promedios de la masterias SAT. ....	8
Figura 3: Histograma de estudiantes testados .....	9
Figura 4: Gráfico de pastel de la variable cualitativa Borough .....	10
Figura 5: Gráfico de barras de frecuencias de la variable cualitativa City. ....	12
Figura 6: Gráfico de correlación #1 .....	13
Figura 7: Gráfico de correlación #2 .....	13
Figura 8: Resultados del ANOVA ONE WAY entre variables promedio SAT y City .....	15

# 1 Introducción

El presente informe se enfoca en el análisis de las puntuaciones SAT en las escuelas del estado de Nueva York. La elección de este conjunto de datos se basa en la relevancia y el impacto que tienen estas puntuaciones en la educación y el futuro de los estudiantes en la ciudad de Nueva York, Estados Unidos. El análisis de estas puntuaciones puede proporcionar información valiosa sobre la calidad de la educación en diferentes escuelas y distritos, lo que puede ser útil para los responsables de la toma de decisiones en el ámbito de la educación.

La base de datos a analizar incluye información detallada sobre las escuelas de Nueva York, el nombre de la escuela, el distrito, ciudad, estado, horario de inicio y fin, matrícula de estudiantes y porcentaje de estudiantes de diferentes grupos étnicos. Además, se proporcionan las puntuaciones promedio de SAT en asignaturas como matemáticas, lectura y escritura, así como el porcentaje de estudiantes que realizaron la prueba.

## **2 Objetivos**

### **2.1 Objetivo General**

Evaluar las puntuaciones SAT en las escuelas de Nueva York y determinar si existen diferencias significativas en las puntuaciones en función de la ubicación geográfica y la diversidad étnica.

### **2.2 Objetivos Específicos**

1. Identificar diferencias en las puntuaciones SAT en función de la ubicación geográfica
2. Investigar la correlación entre la diversidad étnica y las puntuaciones SAT
3. XYZ

### 3 Metodología

Para la realización de este proyecto se utilizó el lenguaje de programación **R**, empleando un conjunto de librerías que facilitaron el análisis de los datos.

De la base de datos original se extrajeron las variables que se consideraron de valor para el análisis, las cuales fueron:

#### **Variables cualitativas:**

- Borough: Distrito de la escuela
- City : Ciudad de la escuela

#### **Variables cuantitativas:**

- Student\_Enrollment: Cantidad de estudiantes matriculados
- Percent\_White: Porcentaje de estudiantes blancos
- Percent\_Black: Porcentaje de estudiantes negros
- Percent\_Asian: Porcentaje de estudiantes asiáticos
- Average\_Score\_SAT\_Math: Promedio de las puntuaciones SAT en matemáticas
- Average\_Score\_SAT\_Reading: Promedio de las puntuaciones SAT en escritura
- Average\_Score\_SAT\_Writing: Promedio de las puntuaciones SAT en lectura
- Percent\_Testes: Porcentaje de estudiantes que realizaron la prueba
- Average\_SAT: Promedio de las puntuaciones SAT

## 4 Variables de la base de Datos

### 4.1 Variables Cuantitativas

#### 4.1.1 Estudiantes Blancos, Negros y Asiáticos

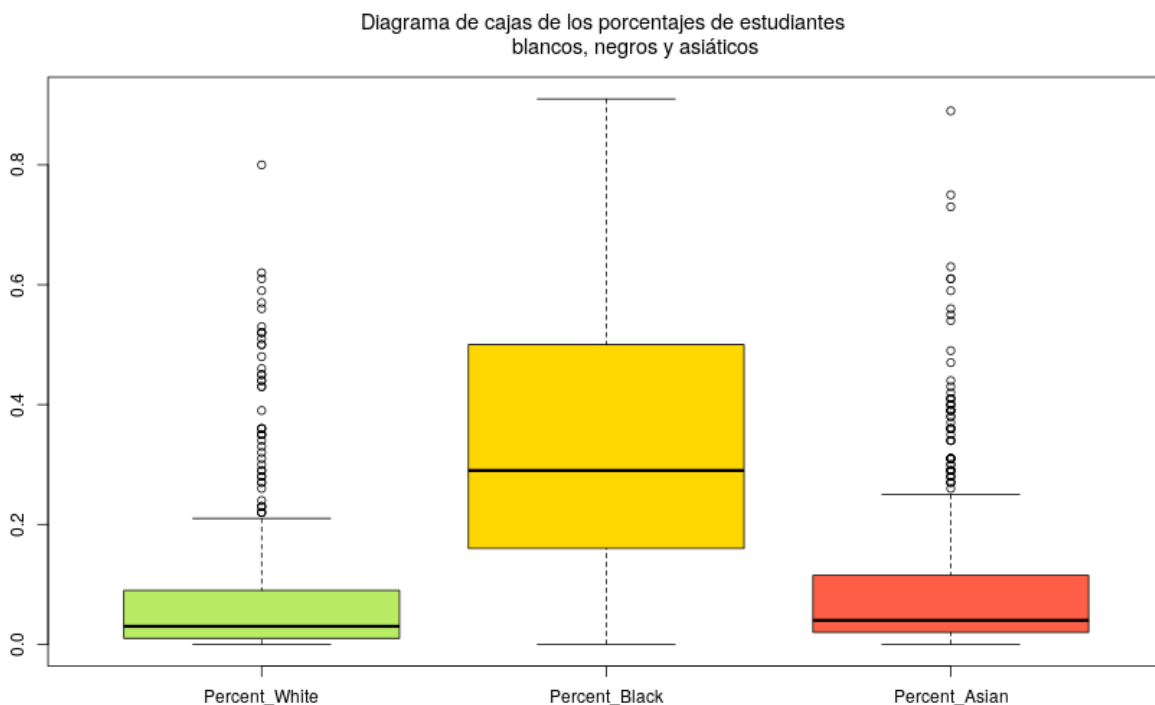


Figura 1: Diagrama de cajas de los porcentajes de estudiantes blancos, negros y asiáticos.

El presente gráfico de cajas representa los porcentajes de estudiantes de etnias blancos, negros y asiáticos. Los estudiantes blancos, representados por la caja verde, tienen un rango de porcentajes que varía entre 0.01 y 0.61, con una mediana alrededor de 0.04 y algunos valores atípicos en el extremo superior. Los estudiantes de etnia negros, representados por la caja amarilla, tienen un rango de porcentajes que varía entre 0.03 y 0.53, con una mediana alrededor de 0.28 y algunos valores atípicos en ambos extremos. Los estudiantes asiáticos, representados por la caja roja, tienen un rango de porcentajes que varía entre 0.0 y 0.89, con una mediana alrededor de 0.09 y varios valores atípicos en el extremo superior.

#### 4.1.2 Promedios de SAT

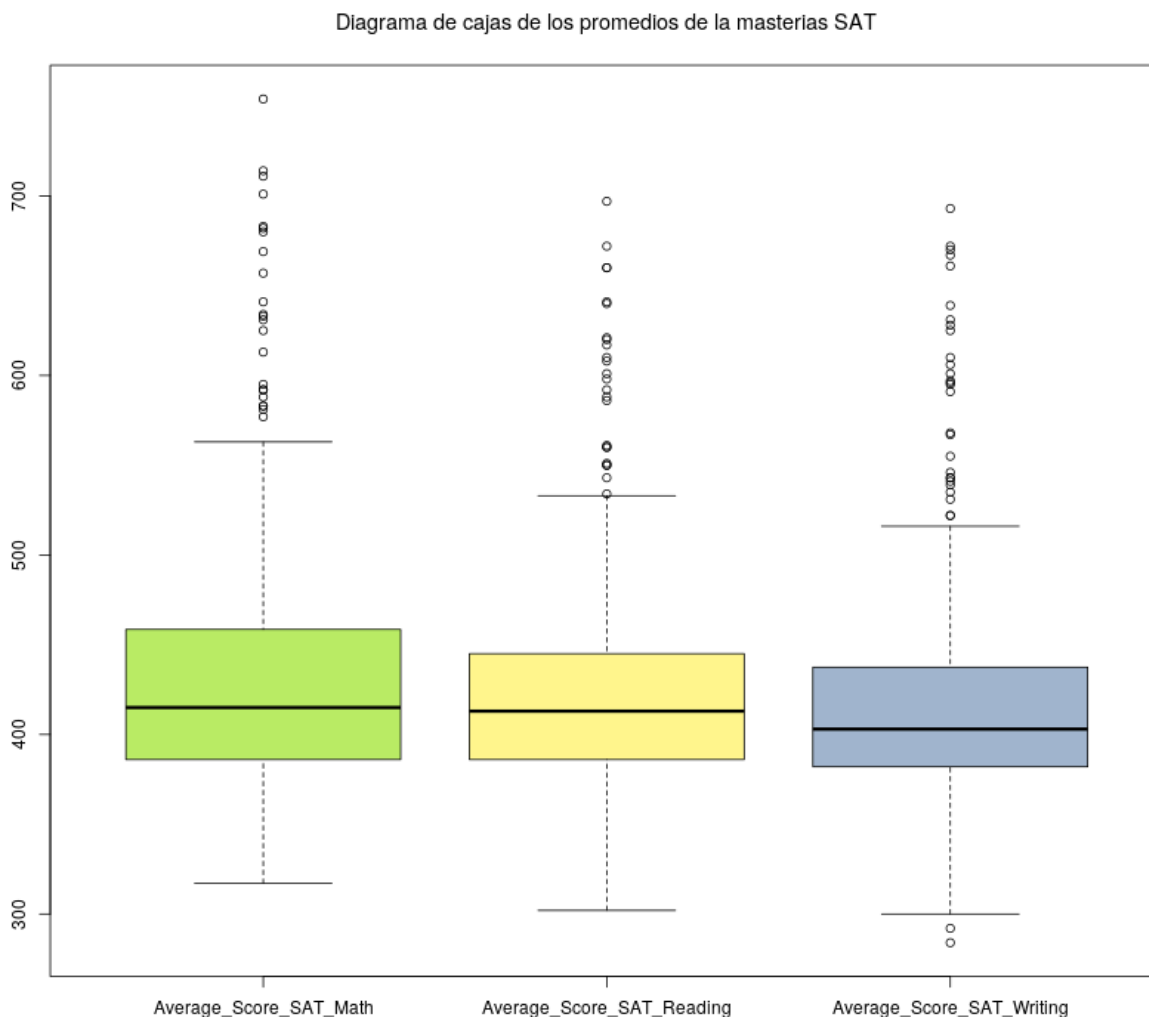


Figura 2: Diagrama de cajas de los promedios de la masterías SAT.

El presente gráfico de cajas que representa los promedios de las puntuaciones SAT en Matemáticas, Lectura y Escritura. En Matemáticas, el 25% de los estudiantes obtuvo una puntuación de 386 o menos, el 50% obtuvo una puntuación de 415 o menos, y el 75% obtuvo una puntuación de 459 o menos. En Lectura, los cuartiles son 386, 413 y 445 respectivamente, lo que indica que las puntuaciones en lectura son generalmente más altas que en matemáticas. En Escritura, los valores son 382, 403 y 438 para cada uno de los tres cuartiles respectivamente, lo que sugiere que las puntuaciones en escritura son generalmente más bajas que en las otras dos asignaturas. Los puntos por encima de cada caja representan posibles valores atípicos, es decir, puntuaciones que se desvían significativamente del resto de los datos.



### 4.1.3 Estudiantes Testados

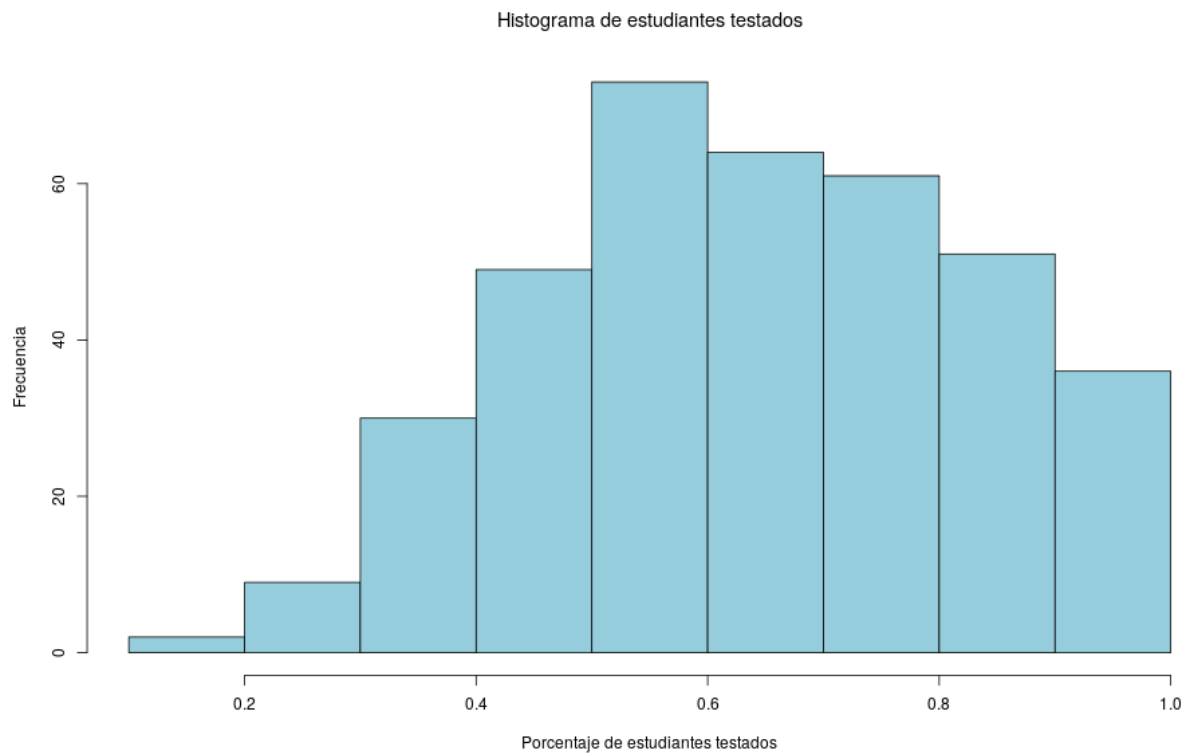


Figura 3: Histograma de estudiantes testados

La gráfica de estilo histograma anterior representa la distribución de los porcentajes de estudiantes que realizaron la prueba **SAT**. La mayor frecuencia de estudiantes testados se encuentra en el rango de 0.5 a 0.7. Con respecto a la distribución de los datos, se encuentra una distribución no uniforme, con un comportamiento irregular.

## 4.2 Variables Cualitativas

### 4.2.1 Borough

Gráfico de pastel de la variable cualitativa Borough

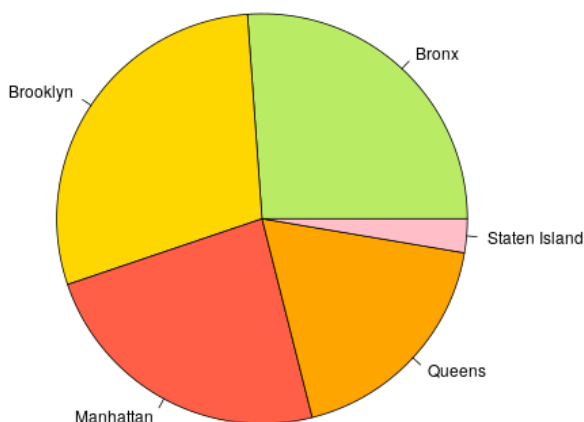


Figura 4: Gráfico de pastel de la variable cualitativa Borough

El gráfico de pastel presentado muestra una representación visual detallada de la distribución porcentual de la población en los cinco distritos de la ciudad de Nueva York, que son Brooklyn, Bronx, Manhattan, Queens y Staten

Island. Cada segmento del gráfico de pastel representa un distrito y su tamaño es proporcional al porcentaje de la población que reside en ese distrito. Brooklyn, con el 29,07% de la población, tiene la mayor proporción, lo que se refleja en el segmento más grande del gráfico de pastel. Esto indica que Brooklyn es el distrito más poblado de la ciudad de Nueva York. El Bronx, con el 26,13% de la población, tiene el segundo segmento más grande del gráfico de pastel, lo que indica que también es un distrito densamente poblado. Manhattan, tiene el tercer segmento más grande en el gráfico de pastel, representando el 23,73% de la población. Queens, tiene el cuarto segmento más grande en el gráfico de pastel, representando el 18,40% de la población. Staten Island, tiene el segmento más pequeño en el gráfico de pastel, representando solo el 2,67% de la población. Esto indica que Staten Island tiene la

menor densidad de población entre los cinco distritos. La suma total de las proporciones es del 100,00%, lo que indica que todos los distritos de la ciudad de Nueva York están representados en este gráfico.

### 4.2.2 City

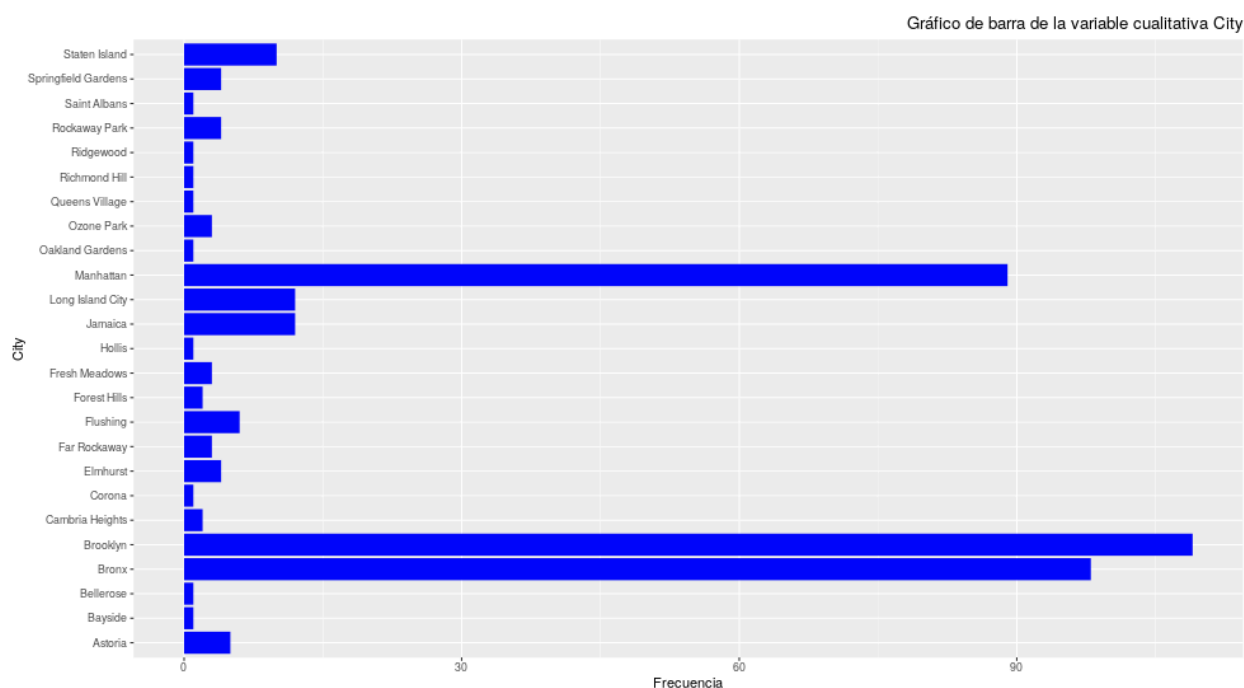


Figura 5: Gráfico de barras de frecuencias de la variable cualitativa City.

El presente diagrama barras presenta la distribución de las ciudades en donde habitan los estudiantes de la base estudiada, los datos mostraron que Brooklyn encabeza la lista con un total de 109 ciudades. Esto indica que Brooklyn es una región con una gran cantidad de ciudades en su territorio. Le sigue de cerca el Bronx, con 98 ciudades, y Manhattan, con 89 ciudades. Por otro lado, Staten Island alberga 10 ciudades, lo que es significativamente menor en comparación con Brooklyn, Bronx y Manhattan. Esto podría indicar que Staten Island tiene una densidad de ciudades mucho menor. Además, hay varios lugares como Bayside, Bellerose, Corona, Hollis, Oakland Gardens, Queens Village, Richmond Hill, Ridgewood y Saint Albans que tienen solo una ciudad cada uno. Esto podría sugerir que estas áreas son menos urbanizadas o que están compuestas principalmente por una sola ciudad grande.

## 5 Análisis y Resultados

### 5.1 Análisis de Correlación entre el promedio SAT y etnias

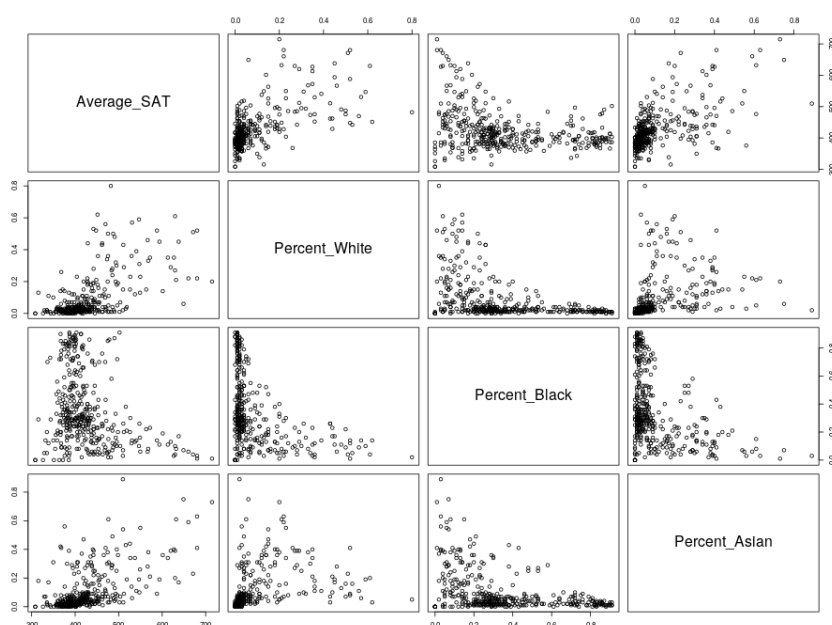


Figura 6: Gráfico de correlación #1

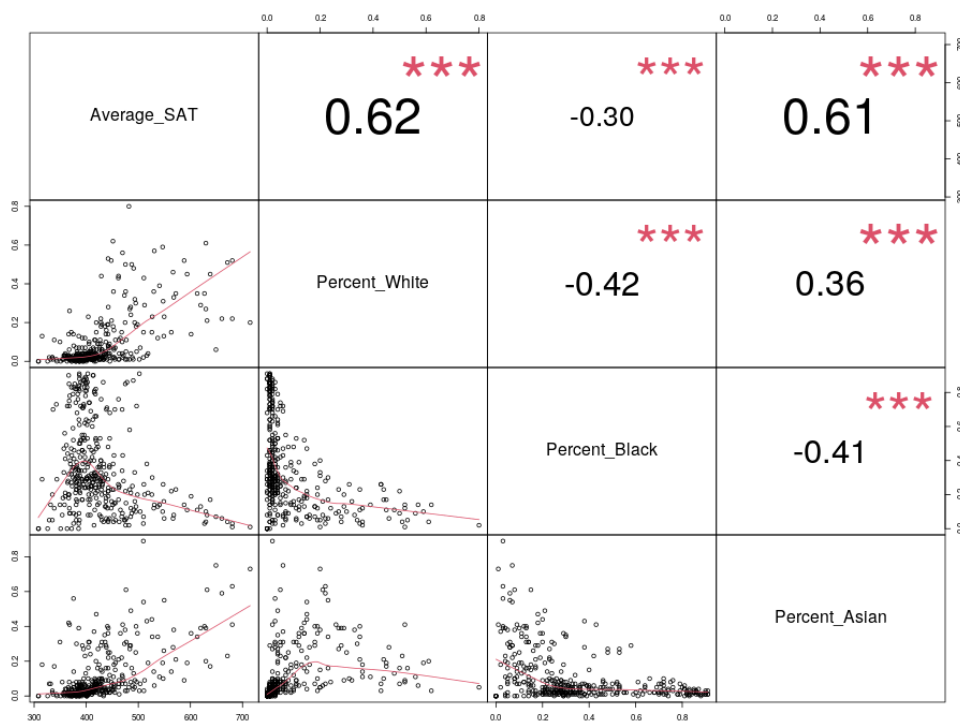


Figura 7: Gráfico de correlación #2

En la Figura 7 se presentan los resultados de la correlación entre las variables Average\_SAT, Percent\_White, Percent\_Black y Percent\_Asian. Se puede observar que la correlación entre Average\_SAT y Percent\_White es positiva con un valor de 0.6234902, lo que indica que existe una relación directa entre estas dos variables. Por otro lado, la correlación entre Average\_SAT y Percent\_Black es negativa con un valor de -0.3048109, lo que sugiere una relación inversa entre ambas variables. Así mismo, se puede apreciar que la correlación entre Average\_SAT y Percent\_Asian es positiva con un valor de 0.6098355, lo que indica una relación directa entre estas dos variables. Por último, la correlación entre Percent\_White y Percent\_Black es negativa con un valor de -0.4220592, mientras que la correlación entre Percent\_White y Percent\_Asian es positiva con un valor de 0.3555783.

Estos resultados muestran que existe una relación entre las variables analizadas y que es importante tener en cuenta su impacto en los resultados del **SAT**.

## 5.2 ANOVA ONE WAY entre variables promedio SAT y City

```
      Df Sum Sq Mean Sq F value Pr(>F)
data$City 24 258367   10765   2.854 1.45e-05 ***
Residuals 350 1320272    3772
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 8: Resultados del ANOVA ONE WAY entre variables promedio SAT y City

La figura anterior muestra el análisis de los resultados del ANOVA en donde se revela una diferencia significativa entre las variables Average\_SAT y City.

Según los datos obtenidos, se puede concluir que la ciudad de residencia de los estudiantes influye de manera significativa en el rendimiento promedio del examen **SAT**. Con un valor de F de 2.854 y un p-valor de 1.45e-05, se puede afirmar con seguridad que existe una relación entre ambas variables. Pese a esto, es importante tener en cuenta que aún hay un alto porcentaje de variabilidad que no puede ser explicado por la ciudad de residencia, lo cual sugiere que existen otros factores que también pueden influir en el rendimiento en el examen SAT.

## 6 Conclusiones

### 6.1 Correlación entre la diversidad étnica y las puntuaciones SAT

En base a los resultados obtenidos, con respecto al objetivo **1**, se extraen las siguientes conclusiones. En primer lugar, se puede afirmar que existe una significativa correlación entre la diversidad étnica y las puntuaciones \*SAT\*. De forma específica, se encontró una correlación positiva entre el porcentaje de estudiantes blancos, asiáticos y las puntuaciones promedio del \*SAT\*. Esto indica que a medida que aumenta el porcentaje de estos grupos étnicos, también lo hacen las puntuaciones del \*SAT\*. Sin embargo, se observó una correlación negativa entre el porcentaje de estudiantes negros y las puntuaciones promedio del \*SAT\*, lo que sugiere que a medida que aumenta el porcentaje de estudiantes negros, las puntuaciones del \*SAT\* tienden a disminuir.

De igual manera, se aprecia una correlación negativa entre el porcentaje de estudiantes blancos y negros, indicando que en las escuelas con un alto porcentaje de estudiantes blancos, hay un porcentaje más bajo de estudiantes negros, y viceversa. También, se ha concluido una correlación positiva entre el porcentaje de estudiantes blancos y asiáticos, lo que sugiere que en las escuelas con un alto porcentaje de estudiantes blancos, también hay un alto porcentaje de estudiantes asiáticos.

A partir del primer objetivo se concluye que se ha identificado una relación entre la diversidad étnica y las puntuaciones del **SAT**. Es posible que existan otros factores no considerados en este estudio que puedan influir en la relación entre la diversidad étnica y las puntuaciones del **SAT**.



## **6.2 Análisis de varianza entre promedios SAT y ubicación**

El análisis de varianza **ANOVA** realizado reveló una diferencia significativa entre los los promedios del SAT y la ciudad de residencia. Según los datos obtenidos, se puede concluir que la ciudad de residencia de los estudiantes influye de manera significativa en el rendimiento promedio del examen SAT. Afirmarmand con seguridad que existe una relación entre los promedios SAT y las residencias.

Es importante tener en cuenta que aún hay un alto porcentaje de variabilidad que no puede ser explicado a partir de la ciudad de residencia. Esto sugiere que existen otros factores que también pueden influir en el rendimiento en el examen SAT. Por lo tanto, aunque la ciudad de residencia es un factor importante, no es el único que determina el rendimiento de los estudiantes en el examen SAT.

Se logra concluir que se ha identificado algunas tendencias y patrones en las puntuaciones SAT en función de la ubicación geográfica. Sin embargo, es posible que existan otros factores no considerados en este estudio que puedan influir en las puntuaciones SAT.

## **7 Anexos**

### **7.1 Código Fuente**

El código fuente de este proyecto se encuentra disponible en el siguiente enlace a GitHub

`<https://github.com/anntnzb/estg1034-proy>.`

## **Bibliografía**