IMPACT OF PARENT'S EDUCATIONAL BACKGROUND ON CHILD'S ACADEMIC ABILITIES

Prepared by: Kiran Neupane

JUNE 28, 2023

THANKOT, KATHMANDU
Kiran10neupane@gmail.com

Table of Contents

Introduction

The purpose of this report is to analyze the impact of Parent's Educational Background on Child's Academic Abilities using a dataset obtained from a public school. The study will focus on two key parameters: Parent's Education and Child's Cumulative Test Score. The data will be cleaned, visualized and analyzed using Python programming language and its associated libraries, including pandas, NumPy and Plotly.

Dataset Description

The data set contains more than 30K entries which are divided into 15 categories as follows:

Gender: Gender of the student (male/female)

EthnicGroup: Ethnic group of the student (group A to E)

ParentEducation: Parent(s) education background (from some_highschool to master's degree)

LunchType: School lunch type (standard or free/reduced)

TestPrep: Test preparation course followed (completed or none)

ParentMaritalStatus: Parent(s) marital status (married/single/widowed/divorced)

PracticeSport: How often the student practice sport (never/sometimes/regularly))

IsFirstChild: If the child is first child in the family or not (yes/no)

NrSiblings: Number of siblings the student has (0 to 7)

TransportMeans: Means of transport to school (schoolbus/private)

WklyStudyHours: Weekly self-study hours (less that 5hrs; between 5 and 10hrs; more than 10hrs)

MathScore: math test score(0-100)

ReadingScore: reading test score(0-100)

WritingScore: writing test score(0-100)

***CumulativeTestScore:** Average of MathScore, ReadingScore and WritingScore(0-100)

*New column **CumulativeTestScore** was manually concatenated in the dataset

Data Cleaning

1. Finding Null Values in the dataset

```
df.isnull().sum()
```

```
Gender                   0
EthnicGroup           1840
ParentEduc            1845
LunchType                0
TestPrep              1830
ParentMaritalStatus   1190
PracticeSport          631
IsFirstChild           904
NrSiblings            1572
TransportMeans        3134
WklyStudyHours         955
MathScore                0
ReadingScore             0
WritingScore             0
dtype: int64
```

2. Replacing Null Values in the dataset

```
cat = (df.dtypes == 'object')
cat_obj = list(cat[cat].index)
for i in cat_obj:
    df[i] = df[i].fillna(method = 'bfill')
```

```
num = (df.dtypes != 'object')
num_obj = list(num[num].index)
for i in num_obj:
    df[i] = df[i].fillna(df[i].mean())
```

3. Calculating CumulativeTestScore

```
# Combining Maths,Reading and Writing Scores for a cumulative average "Test Score"
df["CumulativeTestScore"]= df[["MathScore","ReadingScore","WritingScore"]].mean(axis=1).astype(int)
print(df.head())
```

```
   Gender EthnicGroup          ParentEduc     LunchType TestPrep  \
0  female     group C   bachelor's degree      standard     none
1  female     group C        some college      standard     none
2  female     group B     master's degree      standard     none
3    male     group A  associate's degree  free/reduced     none
4    male     group C        some college      standard     none

  ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings TransportMeans  \
0             married     regularly          yes         3.0     school_bus
1             married     sometimes          yes         0.0     school_bus
2              single     sometimes          yes         4.0     school_bus
3             married         never           no         1.0     school_bus
4             married     sometimes          yes         0.0     school_bus

  WklyStudyHours  MathScore  ReadingScore  WritingScore  CumulativeTestScore
0            < 5         71            71            74                   72
1        5 - 10         69            90            88                   82
2            < 5         87            93            91                   90
3        5 - 10         45            56            42                   47
4        5 - 10         76            78            75                   76
```

4. Calculating Correlation Matrix

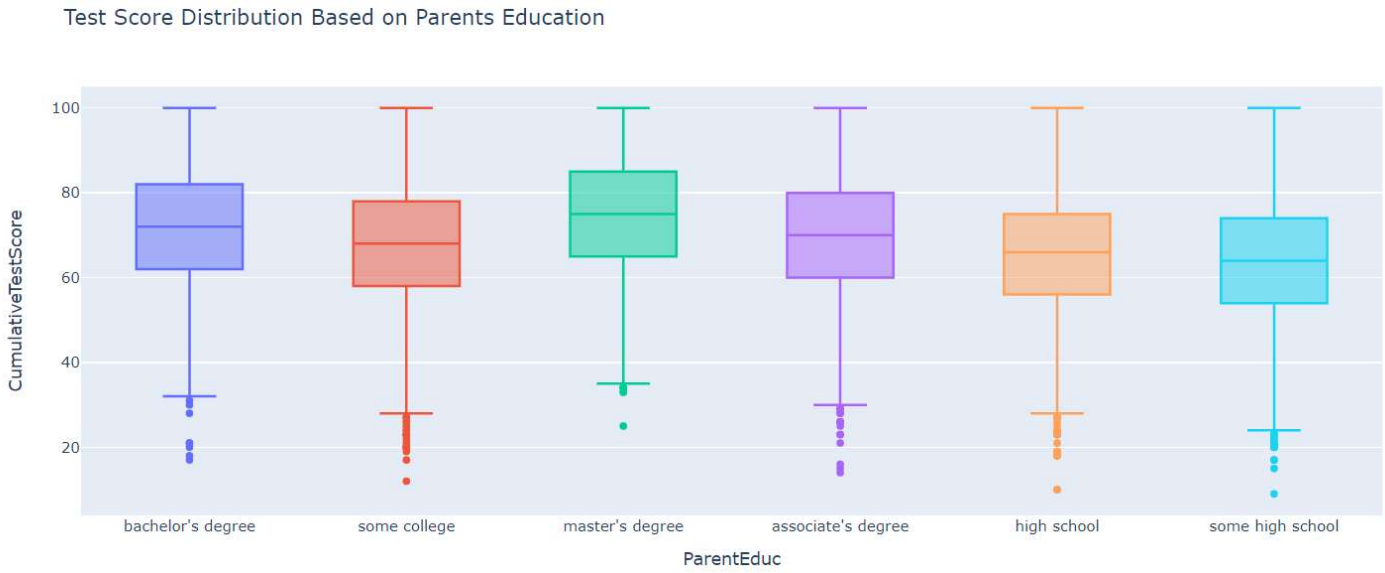
```
# Example dataframe with numerical and encoded categorical columns
data_encoded2= data_encoded[["CumulativeTestScore","associate's degree","bachelor's degree",
                 "high school", "master's degree", "some college","some high school" ]]

# Calculate correlation coefficients
correlation_matrix = data_encoded2.corr()

# Print correlation coefficients
print(correlation_matrix)
```

```
                    CumulativeTestScore  associate's degree  \
CumulativeTestScore            1.000000            0.059129
associate's degree             0.059129            1.000000
bachelor's degree              0.097962           -0.178479
high school                   -0.078112           -0.242149
master's degree                0.120917           -0.134162
some college                  -0.005759           -0.268132
some high school              -0.132815           -0.237893

                    bachelor's degree  high school  master's degree  \
CumulativeTestScore           0.097962    -0.078112         0.120917
associate's degree           -0.178479    -0.242149        -0.134162
bachelor's degree             1.000000    -0.180738        -0.100138
high school                  -0.180738     1.000000        -0.135861
master's degree              -0.100138    -0.135861         1.000000
some college                 -0.200132    -0.271527        -0.150439
some high school             -0.177562    -0.240905        -0.133473

                    some college  some high school
CumulativeTestScore    -0.005759         -0.132815
associate's degree     -0.268132         -0.237893
bachelor's degree      -0.200132         -0.177562
high school            -0.271527         -0.240905
master's degree        -0.150439         -0.133473
some college            1.000000         -0.266755
some high school       -0.266755          1.000000
```

Data Visualization

```
# Plot ParentEducation against CumulativeTestScore.
fig = px.box(df, x = 'ParentEduc', y = 'CumulativeTestScore', color = 'ParentEduc', title = 'Test Score Distribution Based on Parents Education')
fig.show()
```

- Box Plot



Source: Plotly.com

Data Analysis

Impact of Parent's Education Background on Students' Cumulative Test Scores

- A Positive Correlation between ParentEducation and CumulativeTestScore was observed among Parents with a degree (i.e. Master's, Bachelor's or Associate's).
- A Negative Correlation between ParentEducation and CumulativeTestScore was observed among Parents without a degree (i.e. high school, some college or some high school).

Conclusion

These findings underscore the importance of parental education as a significant factor in shaping students' academic success. It highlights the need for educational institutions and policymakers to focus on providing support and resources to students from families with lower educational backgrounds, as they may face additional challenges in achieving academic excellence.

References

[1] "Plotly." Available online: https://plotly.com/. [Accessed: June 28, 2023].

[2] "Kaggle." Available online: https://www.kaggle.com/. [Accessed: June 28, 2023].