

Visualization Pacmann Project

Student Course Performance

arif-gxAl

Oct 8, 2022

Introduction

Dengan meledaknya jumlah data yang bisa didapat dari industry, kebutuhan untuk memahami kondisi bisnis berdasarkan data dengan user-friendly meningkat. Sebagai data analis, dapat mengeksplorasi data dan menunjukkan actionable insight menjadi kemampuan krusial yang harus dimiliki.

Melalui project ini, penulis mencoba untuk melakukan eksplorasi terhadap dataset student courses performance mulai dari menjelaskan *data description*, *research question*, membuat *data visualization* yang sesuai (termasuk *interactive visualization*) dan terakhir ialah membuat *insight* dan rekomendasi yang bisa memberikan masukan bagi para pembacanya. Semoga bermanfaat!

Data Description

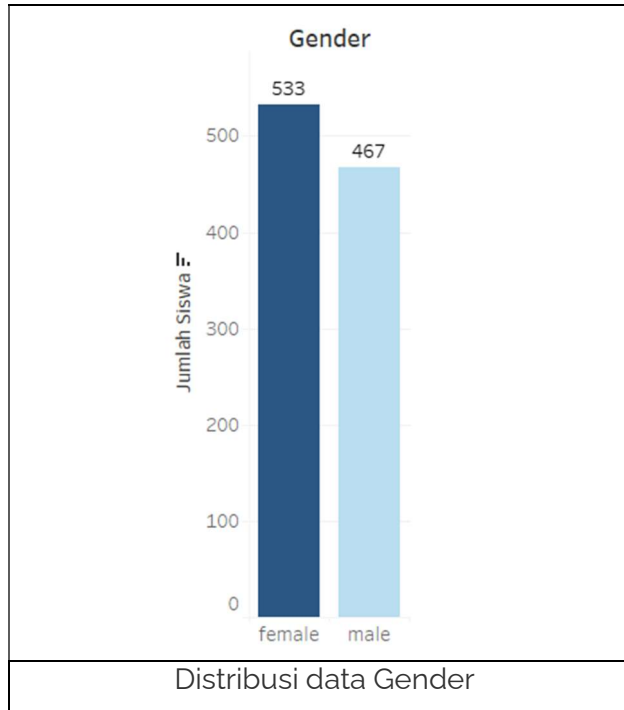
Dataset ini terdiri atas 1000 rows. Masing-masing row mempresentasikan profil siswa dan nilai hasil ujiannya. Keseluruhan dataset ini terdiri atas 8 variables. Berikut ini adalah daftar variable dari dataset tersebut:

1. `gender`: Pria/wanita.
2. `race/ethnicity`: Group yang terdiri dari group A hingga E.
3. `parental level of education`: Detail pendidikan terakhir orang tua siswa mulai dari sekolah menengah atas hingga master's degree.
4. `lunch`: Type jenis makan siang yang dipilih siswa.
5. `test preparation course`: kursus persiapan ujian.
6. `math score`: Nilai matematika dari siswa tersebut.
7. `reading score`: Nilai kemampuan membaca dari siswa tersebut.

8. *writing score*: Nilai kemampuan menulis dari siswa tersebut.

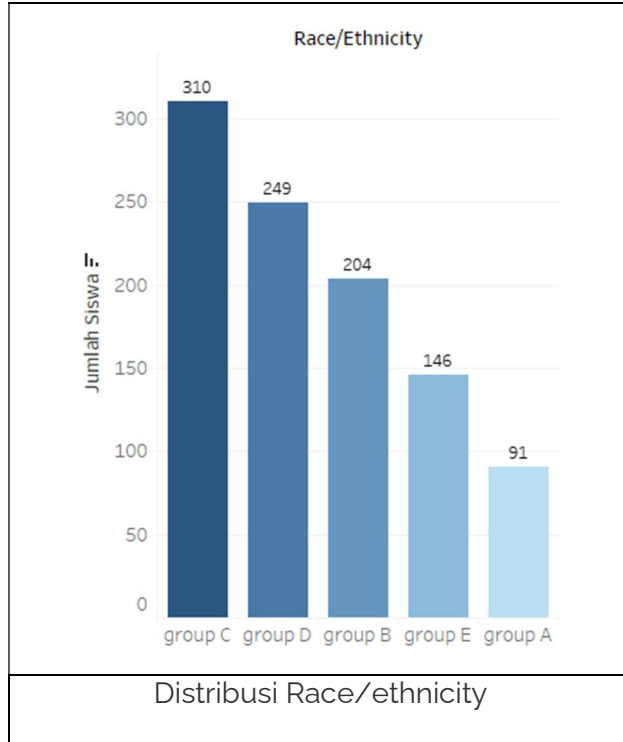
Berikut adalah *descriptive statistics* untuk dataset tersebut.

1. Gender



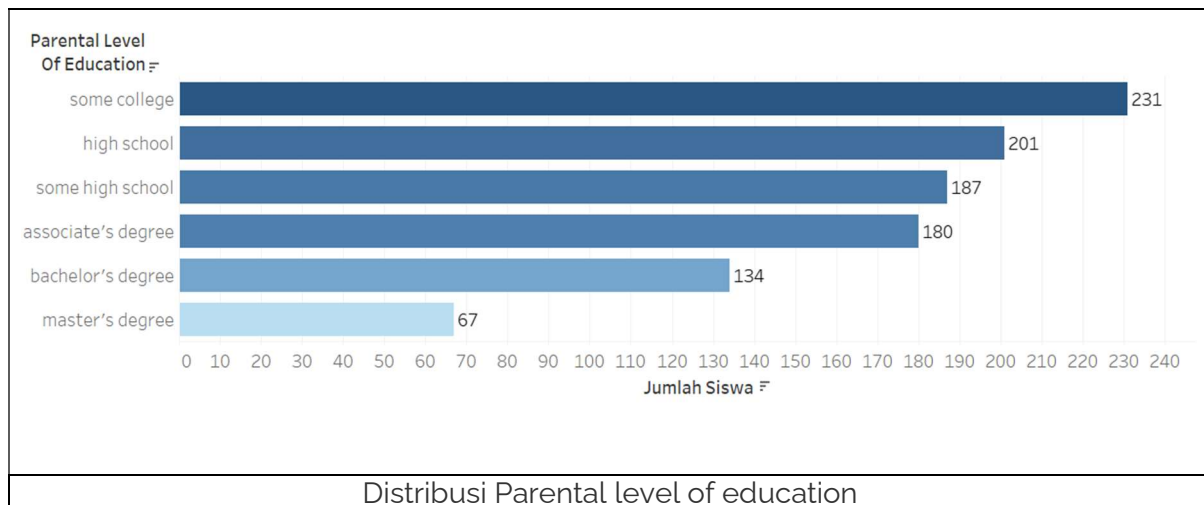
Berdasarkan Bar chart disamping, diketahui bahwa jumlah siswa *female* lebih banyak dibandingkan jumlah siswa *male*. Dimana proporsi siswa *female* sebanyak 53,3% dan siswa *male* 46,7% dari total keseluruhan siswa pada dataset tersebut.

2. Race/ethnicity



Tidak diketahui informasi nama sebenarnya dari variable race/ethnicity tersebut. Namun berdasarkan Bar chart disamping terlihat bahwa *group C* adalah race/ ethnicity terbanyak pada sample tersebut dengan proporsi sebanyak 32%, lalu diikuti oleh *group D* (24,9%), *group B* (20,4%), *group E* (14,6%) dan paling sedikit adalah *group A* (9%).

3. Parental level of education



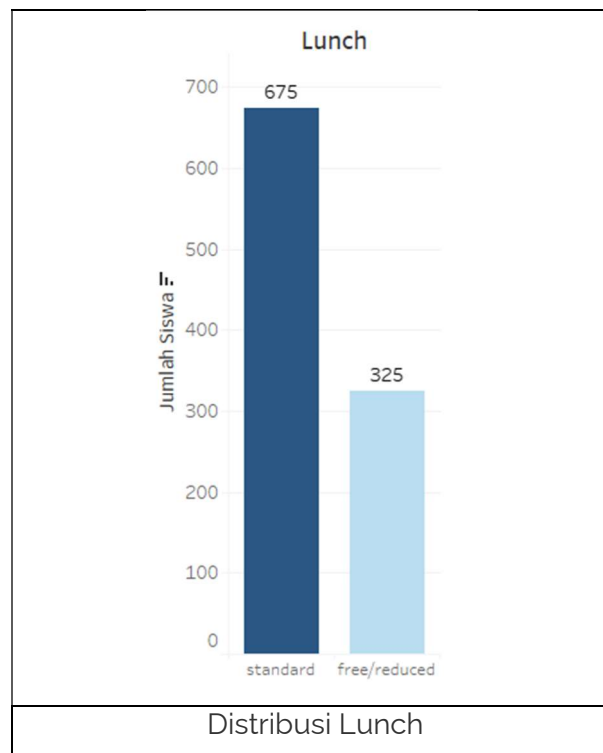
Sekilas tampak proporsi Parent level of education paling banyak adalah *some college* sebesar 23,1%. Namun bila mengacu pada ISCED (*International Standard Classification of Education*), tingkat pendidikan *high school* dan *some high school* adalah sama yaitu *Level 4* dan bila kita

gabungkan maka proporsinya menjadi 38,8%. Kemudian untuk *some college* dan *associate's degree* juga memiliki tingkat pendidikan yang sama yaitu Level 5 dan bila kita gabungkan maka proporsinya menjadi 41,1%.

Berdasarkan penggabungan berdasarkan tingkat pendidikan diatas maka urutan tertinggi ialah *some-college/associate's degree* (Level 5) dengan porsi sebanyak 41,1%, *high school/some high school* (Level 4) 38,8%, *bachelor's degree* (Level 6) 13,4% dan paling sedikit ialah *master's degree* (Level 7) 6,7%.

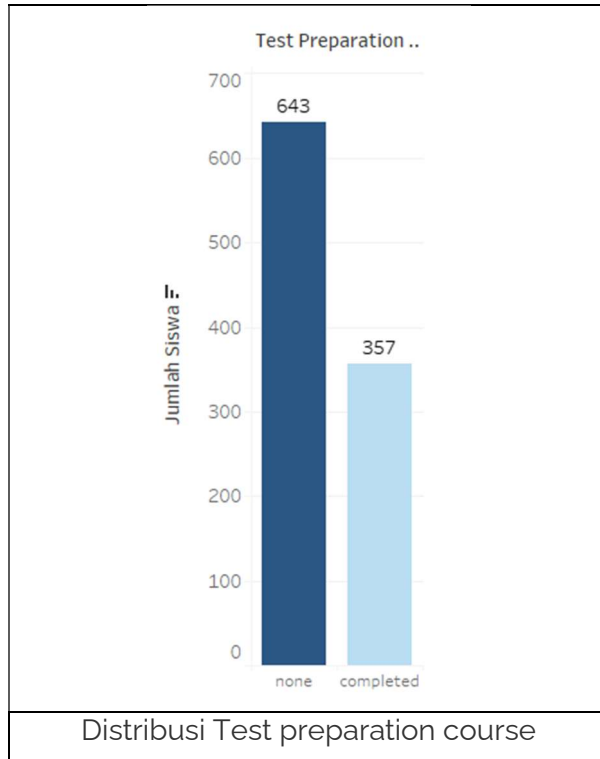
Namun penulis tetap menggunakan nilai parental level of education yang ada tanpa merubah mengikuti acuan tingkat pendidikan ISCED yang dibuat diatas,

4. Lunch



Pada variable ini penulis mengasumsikan bahwa *lunch* merupakan salah satu persiapan yang dilakukan siswa pada saat melakukan test yang ada. Dimana sebanyak 67% siswa mengambil *lunch standard* dan sisanya adalah sebanyak 32,5% tidak mengambilnya.

5. Test preparation course



Test preparation course ini diikuti oleh 35,7% siswa dan sisanya 64,3% siswa tidak mengikuti test ini.

Notes:

Berdasarkan data dibawah ini :

Score Parameter

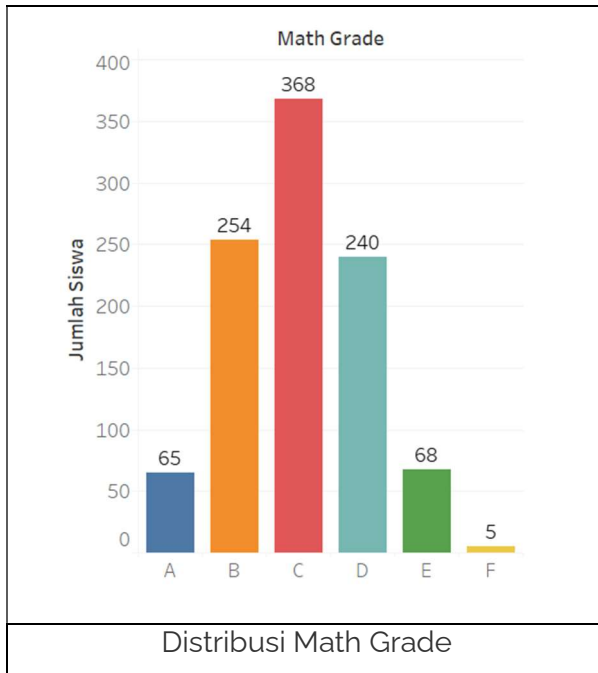
Max. Math Score	100.00
Median Math Score	67.00
Min. Math Score	22.00
Min. Reading Score	29.00
Median Reading Score	70.00
Max. Reading Score	100.00
Min. Writing Score	16.00
Median Writing Score	70.00
Max. Writing Score	100.00

Penulis membagi score pada Math, Reading dan Writing menjadi 6 kategori atau bin 15 dengan informasi sebagai berikut :

Grade	Score
A	90 – 100
B	75 – 90
C	60 – 75

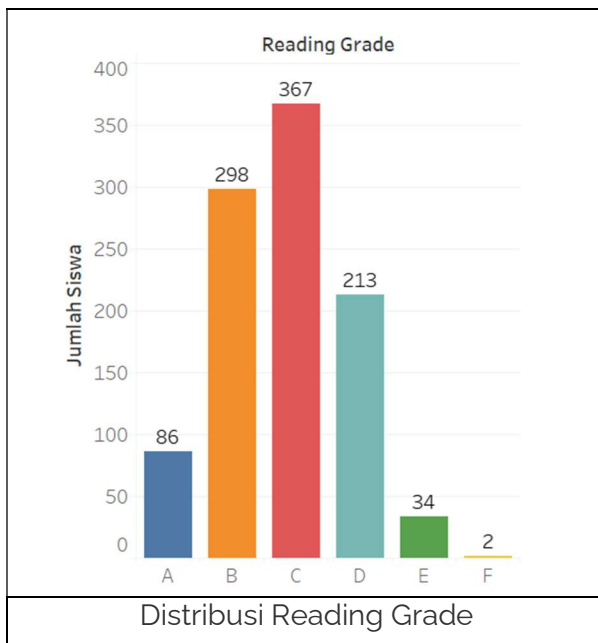
D	45 – 60
E	30 – 45
F	15 - 30

6. Math score



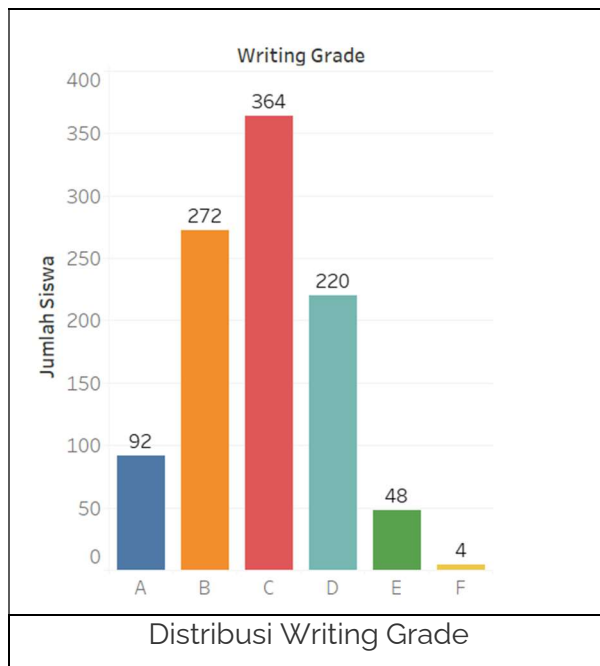
Melalui chart disamping ini, diketahui bahwa Math grade tersebut memiliki distribusi normal. Adapun proporsi terbanyak ialah C (score 60-75) yakni 36,8% lalu diikuti oleh B, D, E, A dan F.

7. Reading score



Melalui chart disamping ini, diketahui bahwa Math grade tersebut memiliki distribusi normal. Adapun proporsi terbanyak ialah C (score 60-75) yakni 36,7% lalu diikuti oleh B, D, E, A dan F.

8. Writing score



Melalui chart disamping ini, diketahui bahwa Math grade tersebut memiliki distribusi normal. Adapun proporsi terbanyak ialah C (score 60-75) yakni 36,4% lalu diikuti oleh B, D, E, A dan F.

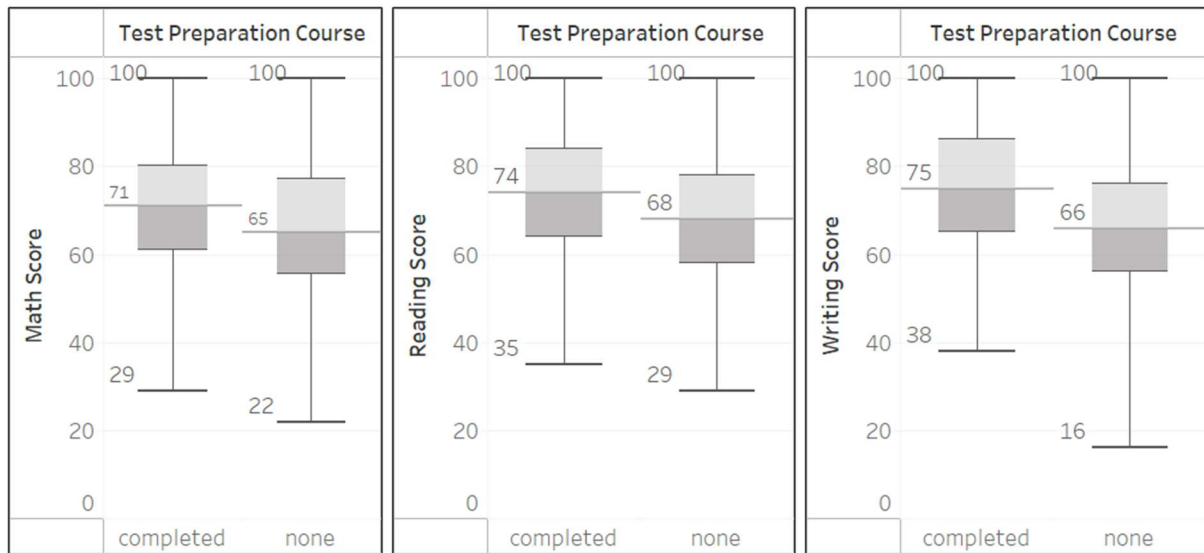
Kesimpulan :

Mayoritas gender dari siswa ialah *female* (53,3%), dengan jumlah race/ethnicity terbanyak adalah *group C* (31%), parental level of education nya ialah *some college* (23,1%), dimana 67,5% siswa mengambil *lunch*, hanya 35,7% yang mengikuti *test preparation course*, dan terakhir ialah grade untuk *Math score*, *Reading score* dan *Writing score* adalah C.

Research Questions

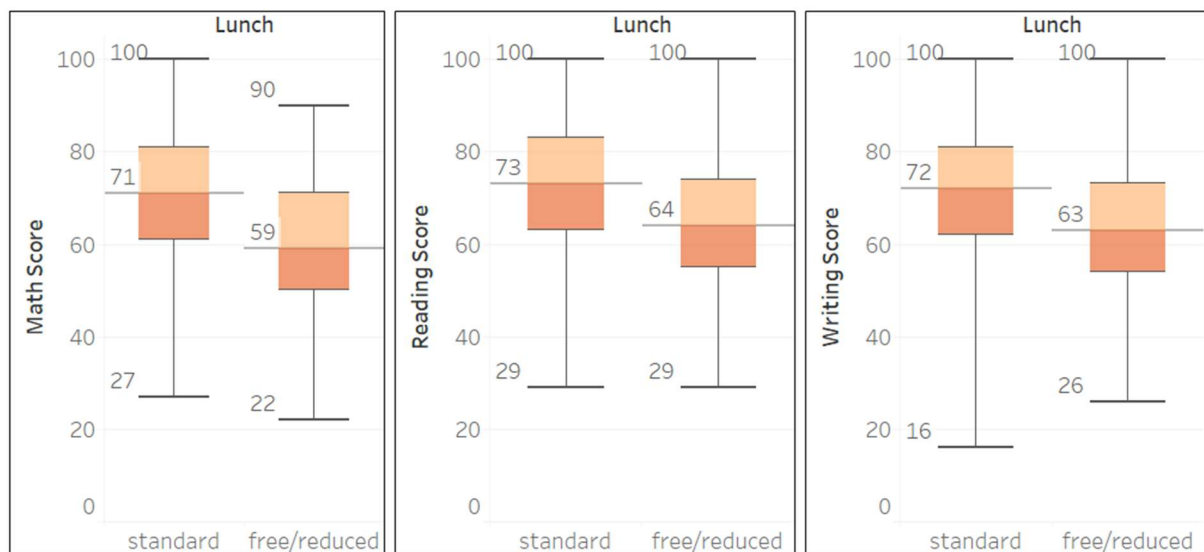
1. Seberapa efektif kah persiapan sebelum test?

- 1.1. Apakah test preparation course mempengaruhi hasil test math, reading, dan writing score?



Ternyata siswa yang melakukan *test preparation course* sebelumnya memiliki score median maupun minimumnya lebih baik dari yang tidak melakukannya.

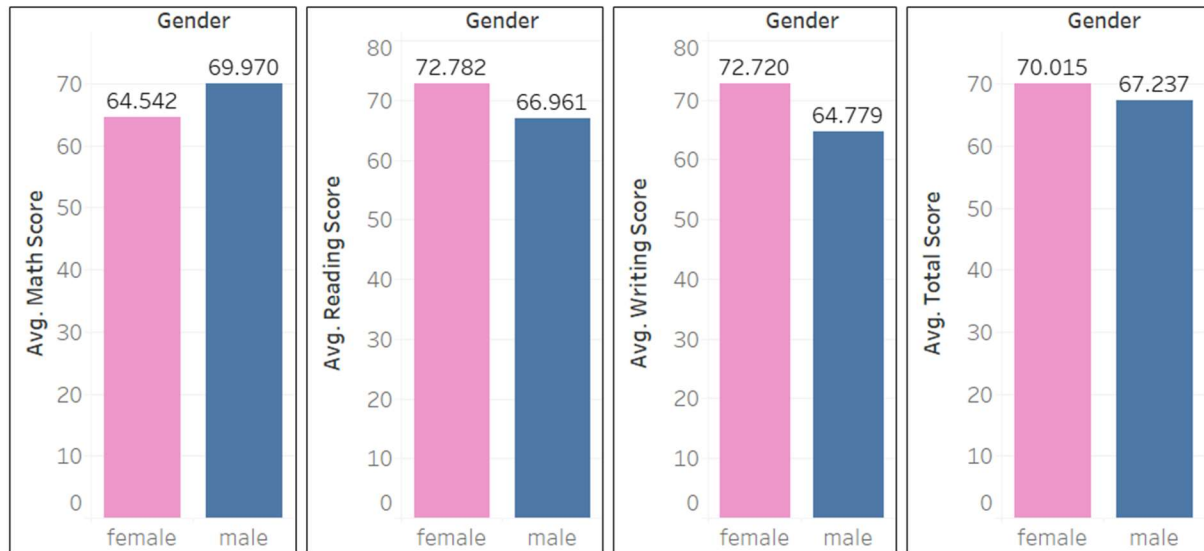
1.2. Apakah lunch mempengaruhi hasil test math, reading, writing?



Hal yang sama terjadi pada siswa yang melakukan *lunch standard* maka median dan minimum score ketiga testnya bernilai lebih dibandingkan dengan siswa *lunch free/reduced* (kecuali minimum score pada writing score, siswa dengan *lunch standard* lebih kecil dibandingkan yang *lunch free/reduced*).

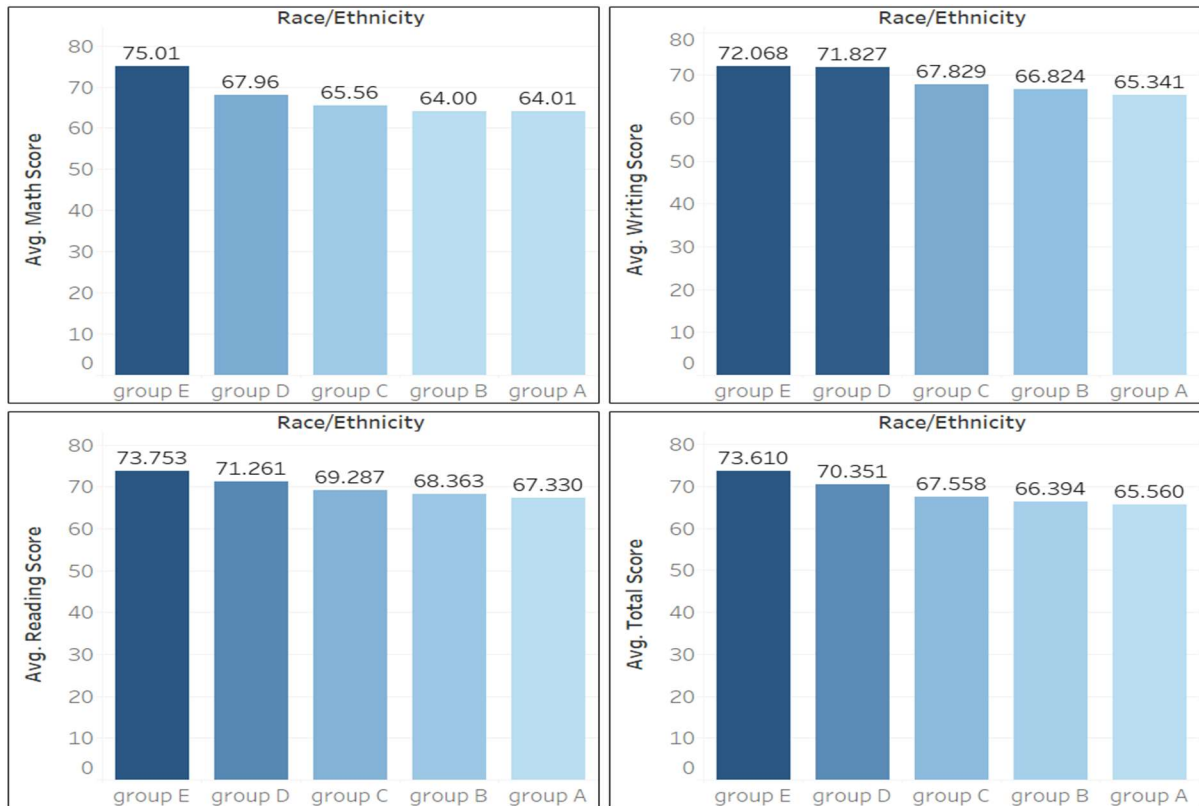
2. Faktor siswa yang berkontribusi terhadap hasil test

2.1. Apakah ada perbedaan hasil test math, reading, writing score, total pada gender yg tertentu ?



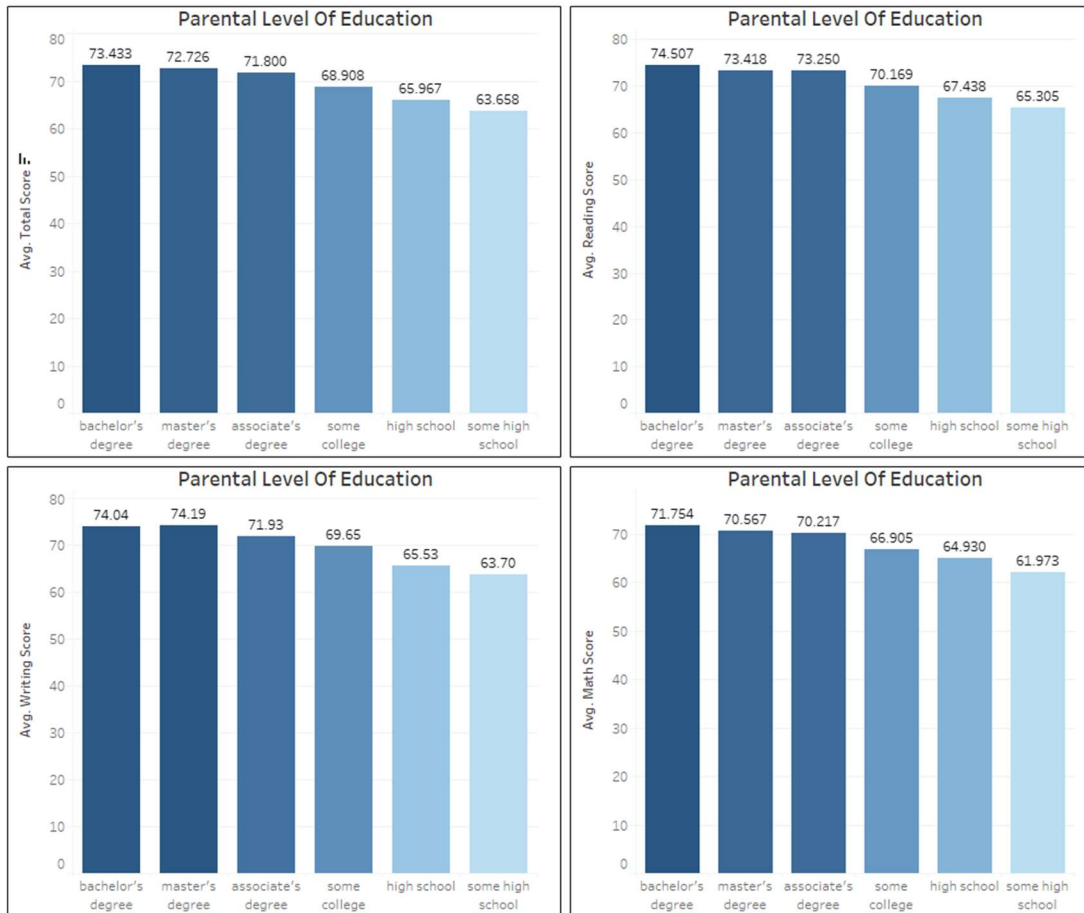
Dari Bar chart diatas bisa dilihat bahwa score tertinggi *math* diraih oleh gender *male*, sedangkan untuk reading, writing dan total score diraih oleh gender *female*. Berdasarkan hal tersebut maka terjawab bahwa adanya perbedaan score masing-masing test dan total score pada gender yang berbeda. Adapun perbedaan diantara keduanya tidak terlalu signifikan.

2.2. Apakah race/ethnicity mempengaruhi hasil test math, reading, writing dan total score?



Bar chart diatas memperlihatkan bahwa score tertinggi untuk masing-masing test dan total score diraih oleh race/ ethnicity *group E*.

- 2.3. Apakah Parental level of education memiliki pengaruh terhadap hasil test math, reading, writing dan total score siswa?



Bar chart diatas memperlihatkan bahwa Parental level of education mempunyai pengaruh pada semua hasil testing yang ada. Dimana orang tua dengan latar belakang pendidikan *bachelor's degree* memiliki score rata tertinggi diatara parental level of education lainnya.

Kecil sekali perbedaan rata-rata score masing-masing test dan rata-rata total score keseluruhan antara orang tua yang berpendidikan *bachelor's degree* dengan *master degree*. Sedangkan nilai rata-rata score terendah ada pada *some high school* dan *high school*, dimana rata-rata score siswa berbanding lurus dengan *parental level of educationnya*.

Conclusion

Berdasarkan hasil EDA yang dilakukan penulis maka disimpulkan bahwa :

1. Parental level of education berbanding lurus dengan rata-rata score yang didapat oleh siswa. Semakin tinggi tingkat parental level of education maka semakin tinggi pula rata-rata score siswanya,
2. Siswa yang mengambil *lunch standard* memiliki nilai median score yang lebih baik daripada yang mengambil *lunch free/reduced*.
3. Siswa yang melakukan *test preparation course* memiliki rata-rata score diseluruh testing yang ada. Sehingga dipastikan untuk bahwa test preparation course akan menambah kesempatan untuk rata-rata score yang lebih tinggi dibandingkan yang tidak melakukannya.
4. Ketiga faktor diatas adalah penting menurut penulis, sedangkan *gender* dan *race/ethnicity* sulit untuk dipastikan keterhubungannya walaupun ada perbedaan rata-rata score diantara nilai yang ada pada variable tersebut. Ada baiknya EDA melakukan operasi correlation pada kesempatan selanjutnya.

Secara umum siswa yang mendapat gizi yang baik, persiapan test yang lebih banyak dan background pendidikan orang tuanya akan sangat mempengaruhi hasilnya.

-AB12-

Link dashboard:

<https://public.tableau.com/app/profile/abudiman1201/viz/StudentCoursePerformance-DecriptiveStatistics/DescriptiveStatistics?publish=yes>

<https://public.tableau.com/app/profile/abudiman1201/viz/StudentCoursePerformance-ResearchQuestion1/FinalScore-Lunch-TestPre?publish=yes>

<https://public.tableau.com/app/profile/abudiman1201/viz/StudentCoursePerformance-ResearchQuestion2/DashScore-ParentEdu?publish=yes>

Link video:

<https://youtu.be/7dESiL-OSAO>