**TUGAS   
PROYEK SAINS DATA**

**Analisis Performa Siswa selama ujian menggunakan Power Bi**



**Dosen pengampu :**

***EKA MALA SARI ROCHMAN, S.Kom.M.Kom***

**Disusun oleh :**

Evy Nur Imamah (210411100123)

**PROGRAM STUDI TEKNIK INFORMATIKA**

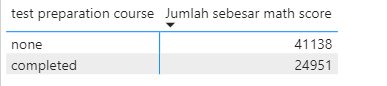
**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

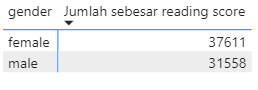
**2023**

1. Cari Data pada Kaggle

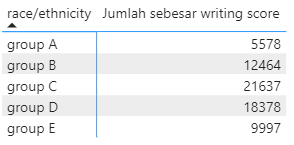
1. Analis Data



Penjelasan : berdasarkan sumbu x (test preparation course) dan sumbu y (math score) dihasilkan besaran presentasenya sebesar 37,5% untuk nilai siswa yang sudah menyelesaikan tes awal ujian matematika sedangkan 62,25% presentase siswa yang tidak menyelesaikan tes awal.

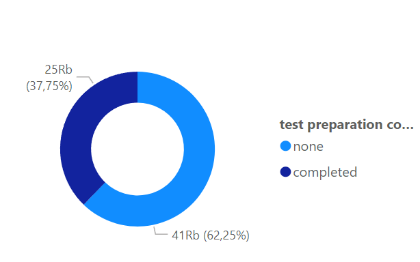


Penjelasan : berdasarkan sumbu x (gender) dan sumbu y (reading score) dihasilkan selisih antar sumbu adalah 2.520 pada kategori free sedangkan untuk kategori standard selisihnya sekitar 3.533.

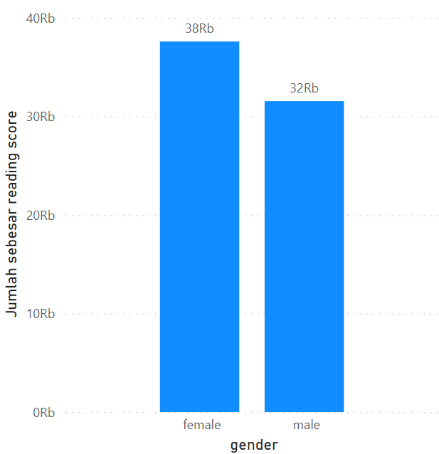


Penjelasan : pada visualiasasi diatas diperoleh dari sumbu x (race/ethnicity) dengan sumbu y (writeng score) diperoleh dari dinas ketenekagaraan seperti master’s degreee dengan rata-rata 559,some senior hs dengan rata-rata 1.242,8 ,senior hs dengan rata-rata 1.253,8

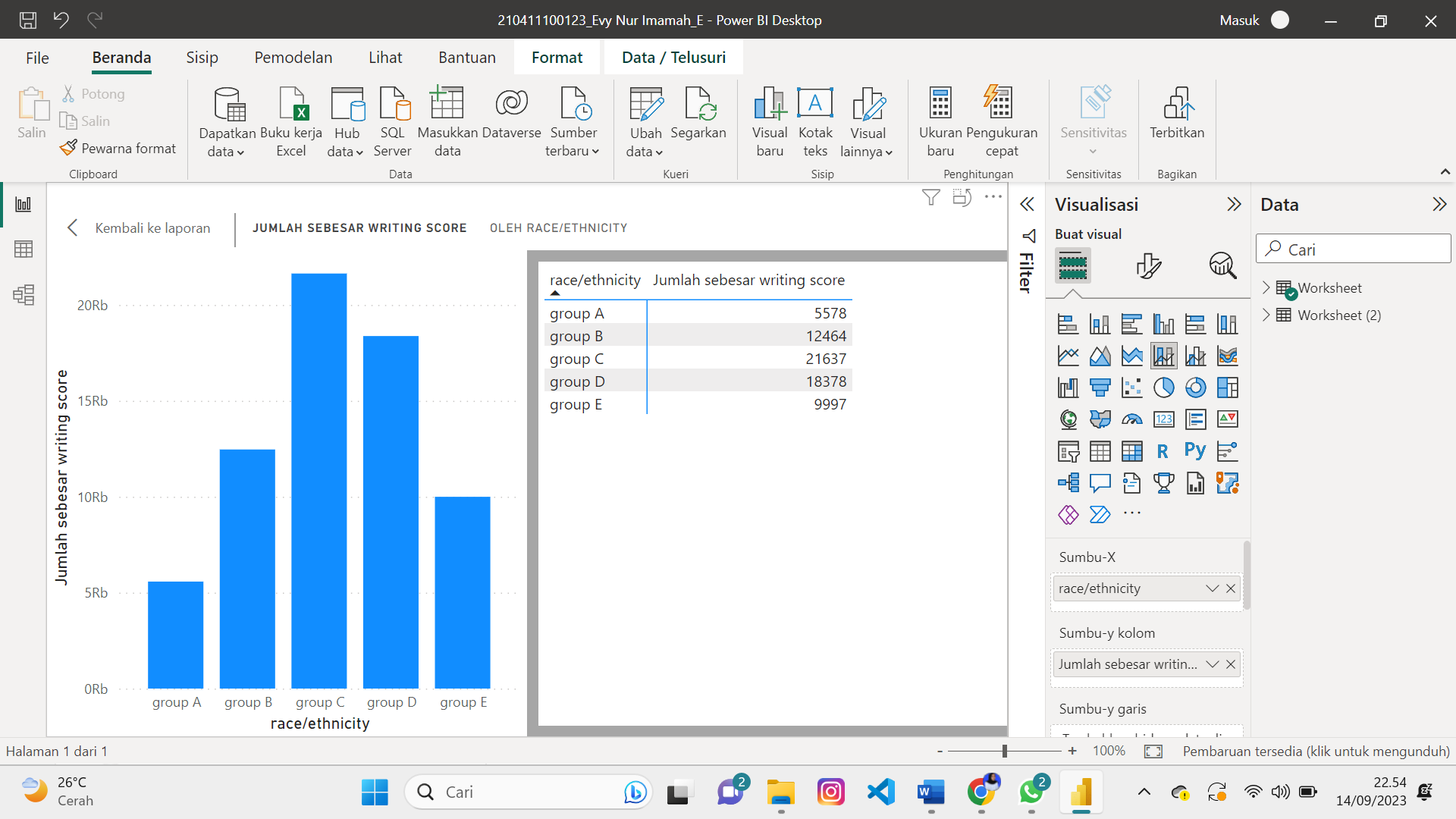
1. Analisis Parameter



Penjelasan : berdasarkan kedua parameter yaitu test preparation course dan math score dihasilkan besaran siswa yang sudah mengerjakan ujian awal matematika sekitar 24.951 lebih sedikit dari besaran siswa yang belum mengerjakan sekitar 41.138.



Penjelasan : berdasarkan kedua parameter yaitu gender dan reading score dihasilkan besaran nilai reading score untuk female sekitar 37.611 sedangkan untuk male sekitar 31.558.



Penjelasan : berdasarkan kedua parameter yaitu race atau etchnicity dan writing score dihasilkan nilai tertinggi ada pada group C sebesar 21.637. rentang nilai writing antara 5.578 sampai 21.637.

2. Cara Mengatasi *Missing Value*

*Missing value* adalah nilai yang hilang atau tidak ada dalam data. Hal ini dapat terjadi karena berbagai alasan seperti kesalahan input, data yang tidak lengkap, atau data yang hilang karena alasan lain. Ketika ada *missing value* dalam data, dapat mempengaruhi hasil analisis data dan membuat kesimpulan yang salah. Ada beberapa cara untuk memperbaiki *missing value* dalam data, seperti yang kamu sebutkan tadi. Berikut adalah penjelasan lebih detail untuk masing-masing cara:

1. Mengisi dengan nilai 0 atau 1: Salah satu cara yang sederhana untuk mengatasi missing value adalah dengan mengisi nilai yang hilang dengan nilai 0 atau 1. Namun, cara ini hanya efektif jika *missing value* terjadi pada variabel biner atau kategorikal.

2. Mengisi dengan nilai rata-rata atau *median*: Cara lain untuk mengatasi *missing value* adalah dengan mengisi nilai yang hilang dengan nilai rata-rata atau median dari variabel tersebut. Hal ini dapat dilakukan jika missing value terjadi pada variabel numerik. Namun, perlu diingat bahwa cara ini dapat mempengaruhi hasil analisis data.

3. Menggunakan *Machine Learning*: Cara yang lebih canggih untuk mengatasi *missing value* adalah dengan menggunakan teknik *Machine Learning* seperti *Regresi* atau *K-Means Clustering*. Metode ini dapat membantu dalam memprediksi nilai yang hilang berdasarkan data yang tersedia. Namun, metode ini memerlukan pengolahan data yang lebih rumit dan memerlukan pengetahuan yang lebih mendalam tentang *Machine Learning*.

3. *Balancing* Data

Untuk menyeimbangkan data yang tidak seimbang, metode SMOTE, metode ini bekerja dengan menambahkan data sintetis pada kelas minoritas, sehingga jumlah data pada kelas minoritas menjadi seimbang dengan kelas mayoritas. Dalam Power BI, kamu bisa dengan mudah menggunakan metode ini dengan membuka editor query, membuat query baru, dan memilih opsi "Transform" dan "SMOTE". Selanjutnya, pilih kolom target dan atur persentase data sintetis yang ingin kamu tambahkan. Dengan cara ini, kamu bisa menyeimbangkan data dengan mudah dan akurat.

Metode SMOTE sangat efektif dalam menyeimbangkan data yang tidak seimbang, terutama pada dataset yang memiliki perbedaan signifikan antara jumlah data pada kelas minoritas dan kelas mayoritas. Dengan menambahkan data sintetis pada kelas minoritas, metode ini memungkinkan kamu untuk memperoleh distribusi data yang lebih seimbang dan akurat. Selain itu, metode SMOTE juga dapat membantu meningkatkan kinerja model dan meminimalkan kesalahan prediksi pada dataset yang tidak seimbang.

Dalam menggunakan metode SMOTE, kamu harus memperhatikan persentase data sintetis yang ingin ditambahkan. Jangan menambahkan terlalu banyak data sintetis, karena hal ini dapat menyebabkan terjadinya *overfitting* (model terlalu fokus pada data pelatihan). Sebaliknya, pastikan bahwa persentase data sintetis yang ditambahkan cukup untuk menyeimbangkan data, tetapi tidak terlalu banyak sehingga dapat mempengaruhi akurasi model.