

KELOMPOK 8

STUDENT PERFORMANCE FACTORS



ANGGOTA



Mahardika Shiddiq
Anshari
2308107010032



Muhammad Nazlul
Ramadhyan
2308107010036



Naufal Farel
Syafilan
2308107010058



Muhammad Syukri
2308107010060

ANALISA FAKTOR PERFORMA SISWA

Sumber dataset:



Link Dataset:

<https://www.kaggle.com/datasets/laingwyn123/student-performance-factors>



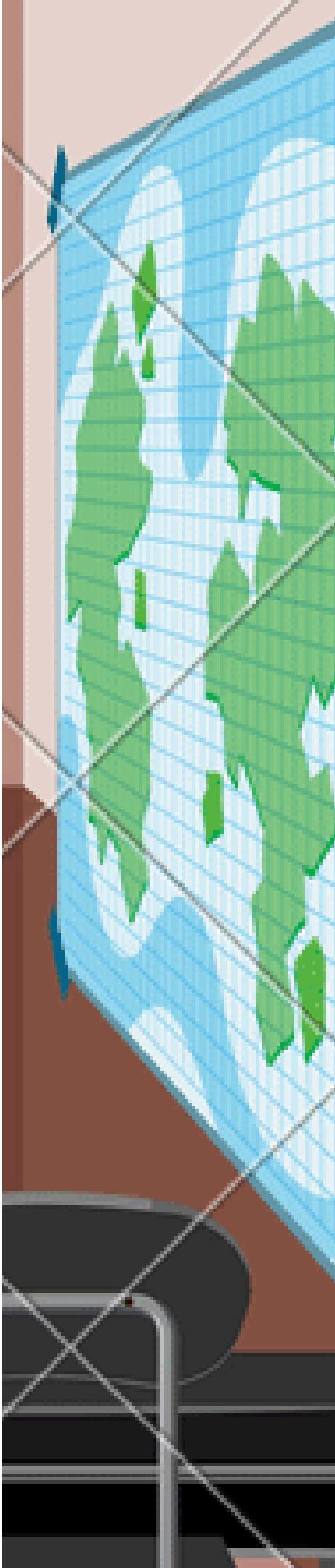


LATAR BELAKANG

- Performa akademik siswa dipengaruhi oleh berbagai faktor internal seperti motivasi, kesehatan mental, dan pola tidur,
- Dipengaruhi pula oleh faktor eksternal seperti kondisi sosial-ekonomi, dukungan keluarga, dan metode pengajaran.
- Untuk memahami hubungan faktor-faktor tersebut, Proyek analisis data diusulkan guna mengidentifikasi pola yang dapat membantu sekolah dan pembuat kebijakan merancang strategi pendidikan yang lebih efektif dan inklusif.



DESKRIPSI DATASET



- Berisi 6.607 data siswa dengan 20 + 2 atribut yang mencerminkan faktor akademik, sosial, dan ekonomi.
- Atribut terdiri dari variabel kuantitatif (misalnya jam belajar, kehadiran, skor sebelumnya) dan kategorikal (seperti jenis kelamin, motivasi, keterlibatan orang tua, kualitas guru).
- Analisis bertujuan untuk mengidentifikasi pola yang memengaruhi kinerja akademik siswa dan mendukung pembuatan kebijakan pendidikan yang lebih efektif.

SAMPLE DATASET

	Hours Studied	Attendance	Parental Involvement	Access to Resources	Extracurricular Activities	Sleep Hours	Previous Scores	Motivation Level	Internet Access	Tutoring Sessions
0	23	84	Low	High	No	7	73	Low	Yes	0
1	19	64	Low	Medium	No	8	59	Low	Yes	2
2	24	98	Medium	Medium	Yes	7	91	Medium	Yes	2
3	29	89	Low	Medium	Yes	8	98	Medium	Yes	1
4	19	92	Medium	Medium	Yes	6	65	Medium	Yes	3
5	19	88	Medium	Medium	Yes	8	89	Medium	Yes	3
6	29	84	Medium	Low	Yes	7	68	Low	Yes	1
7	25	78	Low	High	Yes	6	50	Medium	Yes	1
8	17	94	Medium	High	No	6	80	High	Yes	0
9	23	98	Medium	Medium	Yes	8	71	Medium	Yes	0
10	17	80	Low	High	No	8	88	Medium	No	4
11	17	97	Medium	High	Yes	6	87	Low	Yes	2
12	21	83	Medium	Medium	Yes	8	97	Low	Yes	2
13	9	82	Medium	Medium	Yes	8	72	Medium	Yes	2
14	10	78	Medium	High	Yes	8	74	Medium	Yes	1
15	17	68	Medium	Medium	No	8	70	Medium	Yes	2
16	14	60	Medium	Low	Yes	10	65	Low	Yes	0
17	22	70	Low	Medium	Yes	6	82	Medium	Yes	1
18	15	80	Medium	Medium	Yes	9	91	Low	Yes	3
19	12	75	Medium	High	Yes	7	58	Medium	Yes	3

LANJUTAN SAMPEL DATASET

Family Income	Teacher Quality	School Type	Peer Influence	Physical Activity	Learning Disabilities	Parental Education Level	Distance from Home	Gender	Exam Score
Low	Medium	Public	Positive	3	No	High School	Near	Male	67
Medium	Medium	Public	Negative	4	No	College	Moderate	Female	61
Medium	Medium	Public	Neutral	4	No	Postgraduate	Near	Male	74
Medium	Medium	Public	Negative	4	No	High School	Moderate	Male	71
Medium	High	Public	Neutral	4	No	College	Near	Female	70
Medium	Medium	Public	Positive	3	No	Postgraduate	Near	Male	71
Low	Medium	Private	Neutral	2	No	High School	Moderate	Male	67
High	High	Public	Negative	2	No	High School	Far	Male	66
Medium	Low	Private	Neutral	1	No	College	Near	Male	69
High	High	Public	Positive	5	No	High School	Moderate	Male	72
Medium	High	Private	Neutral	4	No	College	Moderate	Male	68
Low	High	Private	Neutral	2	No	High School	Near	Male	71
Medium	Medium	Public	Positive	4	No	High School	Near	Male	70
Medium	Medium	Private	Positive	3	No	Postgraduate	Near	Male	66
Low	Medium	Private	Neutral	4	No	Postgraduate	Near	Male	65
Medium	Medium	Private	Positive	4	No	High School	Near	Female	64
High	Medium	Private	Positive	3	No	College	Near	Male	60
Low	High	Public	Neutral	3	No	High School	Near	Female	65
Low	Medium	Public	Positive	2	No	College	Moderate	Female	67
Medium	Medium	Private	Positive	4	No	College	Near	Male	66

TIPE DAN RANGE ATRIBUT

- No (kategorikal, 6607)
- Hours_Studied (kuantitatif, range 0-12)
- Attendance (kuantitatif, range 50-100)
- Parental_Involvement (kategorikal, 3 levels: Rendah, Sedang, Tinggi)
- Access_to_Resources (kategorikal, 3 levels: Terbatas, Cukup, Lengkap)
- Extracurricular_Activities (kategorikal, 2 levels: Ya, Tidak)
- Sleep_Hours (kuantitatif, range 3-10)
- Previous_Scores (kuantitatif, range 0-100)
- Motivation_Level (kategorikal, 3 levels: Rendah, Sedang, Tinggi)
- Internet_Access (kategorikal, 2 levels: Ada, Tidak ada)
- Tutoring_Sessions (kuantitatif, range 0-10)
- Family_Income (kategorikal, 3 levels: Rendah, Sedang, Tinggi)
- Teacher_Quality (kategorikal, 3 levels: Rendah, Sedang, Tinggi)
- School_Type (kategorikal, 2 levels: Negeri, Swasta)
- Peer_Influence (kategorikal, 3 levels: Negatif, Netral, Positif)
- Physical_Activity (kuantitatif, range 0-7)
- Learning_Disabilities (kategorikal, 2 levels: Ya, Tidak)
- Parental_Education_Level (kategorikal, 4 levels: Tidak Tamat, SD/Sederajat, SMP/Sederajat, SMA/Ke atas)
- Distance_from_Home (kuantitatif, range 1-50)
- Gender (kategorikal, 2 levels: Laki-laki, Perempuan)
- Exam_Score (kuantitatif, range 55-100)
- Score_Category (kategorikal, 3 levels: Rendah, Sedang, Tinggi)

TUJUAN VISUALISASI

Visualisasi dilakukan untuk melakukan task berupa discover attributes (action target) guna menemukan informasi dari data yang memberikan gambaran akan keterkaitan suatu atribut dalam memengaruhi performa akademik siswa.



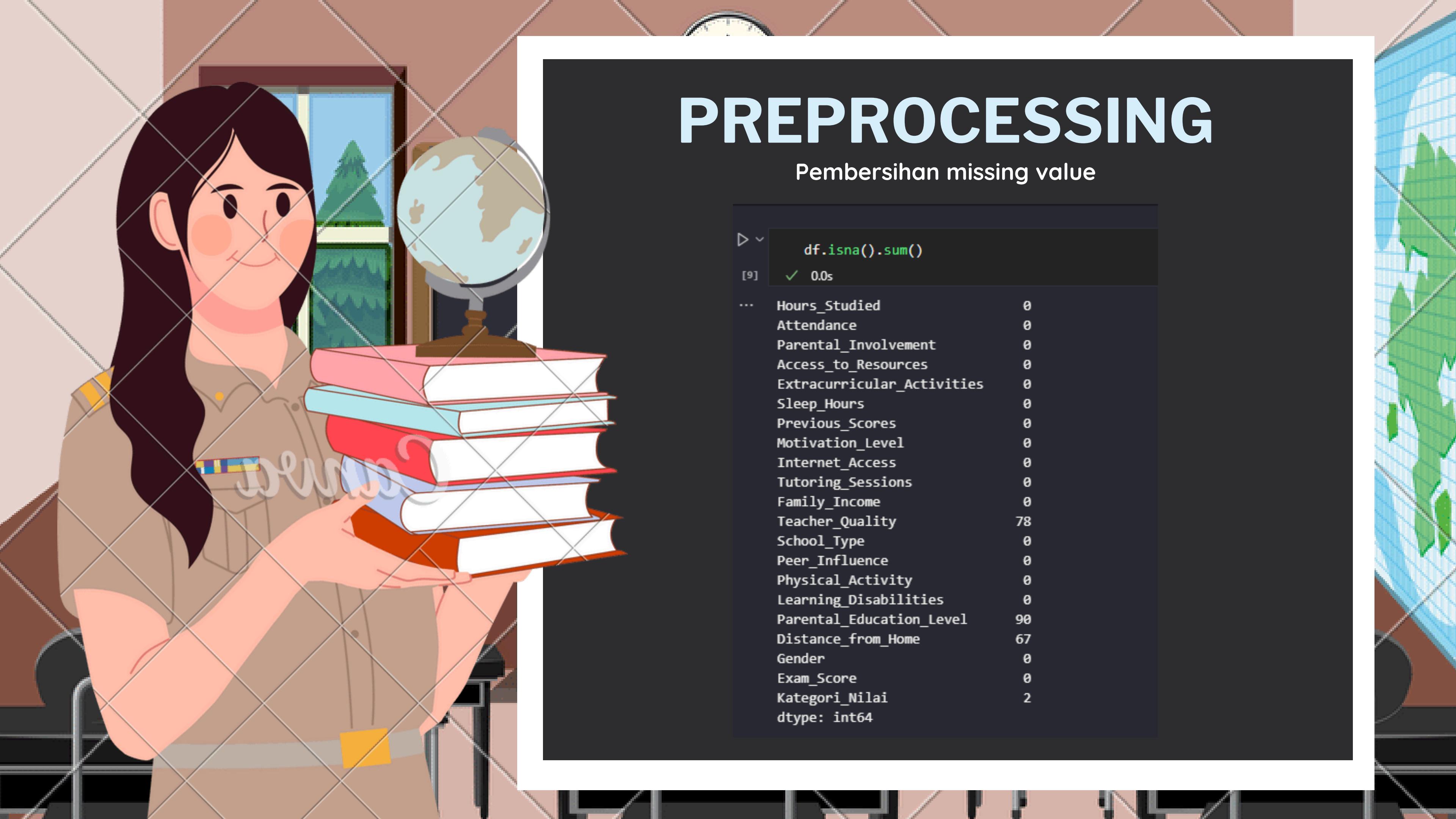
TASK

1. MENINJAU HUBUNGAN ANTARA JUMLAH JAM BELAJAR SISWA DENGAN NILAI UJIAN AKHIR.
2. MENGEVALUASI APAKAH ADA PERBEDAAN PERFORMA AKADEMIK BERDASARKAN TINGKAT KETERLIBATAN ORANG TUA DALAM PENDIDIKAN SISWA.
3. MEMBANDINGKAN PERSENTASE KETINGGIAN NILAI BERDASARKAN JENIS SEKOLAH (NEGERI ATAU SWASTA).
4. MENINJAU DISTRIBUSI TINGKAT MOTIVASI BELAJAR SISWA DAN HUBUNGANNYA DENGAN PERFORMA AKADEMIK.
5. MENGANALISIS APAKAH FAKTOR SOSIAL SEPERTI PENGARUH TEMAN SEBAYA MEMILIKI DAMPAK SIGNIFIKAN TERHADAP NILAI UJIAN AKHIR SISWA.
6. MENILAI HUBUNGAN ANTARA TINGKAT PENDIDIKAN ORANG TUA DAN KONDISI KEUANGAN KELUARGA.
7. MENGANALISIS PENGARUH KEUANGAN KELUARGA DENGAN PERFORMA BELAJAR SISWA.
8. HUBUNGAN ANTARA JAM BELAJAR DENGAN TINGKAT KEHADIRAN SISWA



LANGKAH ANALISIS

1. Eksplorasi Awal Dataset
 2. Pembersihan Data
 3. Pemilihan Atribut Penting
 4. Desain Visualisasi
 5. Implementasi Visualisasi
 6. Analisis dan Interpretasi
- 



PREPROCESSING

Pembersihan missing value

```
df.isna().sum()  
[9] ✓ 0.0s  
... Hours_Studied 0  
Attendance 0  
Parental_Involvement 0  
Access_to_Resources 0  
Extracurricular_Activities 0  
Sleep_Hours 0  
Previous_Scores 0  
Motivation_Level 0  
Internet_Access 0  
Tutoring_Sessions 0  
Family_Income 0  
Teacher_Quality 78  
School_Type 0  
Peer_Influence 0  
Physical_Activity 0  
Learning_Disabilities 0  
Parental_Education_Level 90  
Distance_from_Home 67  
Gender 0  
Exam_Score 0  
Kategori_Nilai 2  
dtype: int64
```

PREPROCESSING

kode untuk membersihkan missing value

```
# Menghapus baris yang memiliki missing values  
df_cleaned = df.dropna(subset=['Parental_Education_Level', 'Distance_from_Home', 'Teacher_Quality', 'Kategori_Nilai'])  
  
# Menampilkan data setelah penghapusan missing values  
df_cleaned.head()
```

CodelImage

PREPROCESSING

Hasil pembersihan missing value

```
[18]: df.isna().sum()  
      ✓ 0.0s  
  
... Hours_Studied           0  
Attendance                  0  
Parental_Involvement        0  
Access_to_Resources          0  
Extracurricular_Activities   0  
Sleep_Hours                 0  
Previous_Scores              0  
Motivation_Level             0  
Internet_Access              0  
Tutoring_Sessions            0  
Family_Income                0  
Teacher_Quality               0  
School_Type                  0  
Peer_Influence                0  
Physical_Activity             0  
Learning_Disabilities         0  
Parental_Education_Level      0  
Distance_from_Home            0  
Gender                        0  
Exam_Score                    0  
Kategori_Nilai                0  
dtype: int64
```



PREPROCESSING

Mencari data duplikat

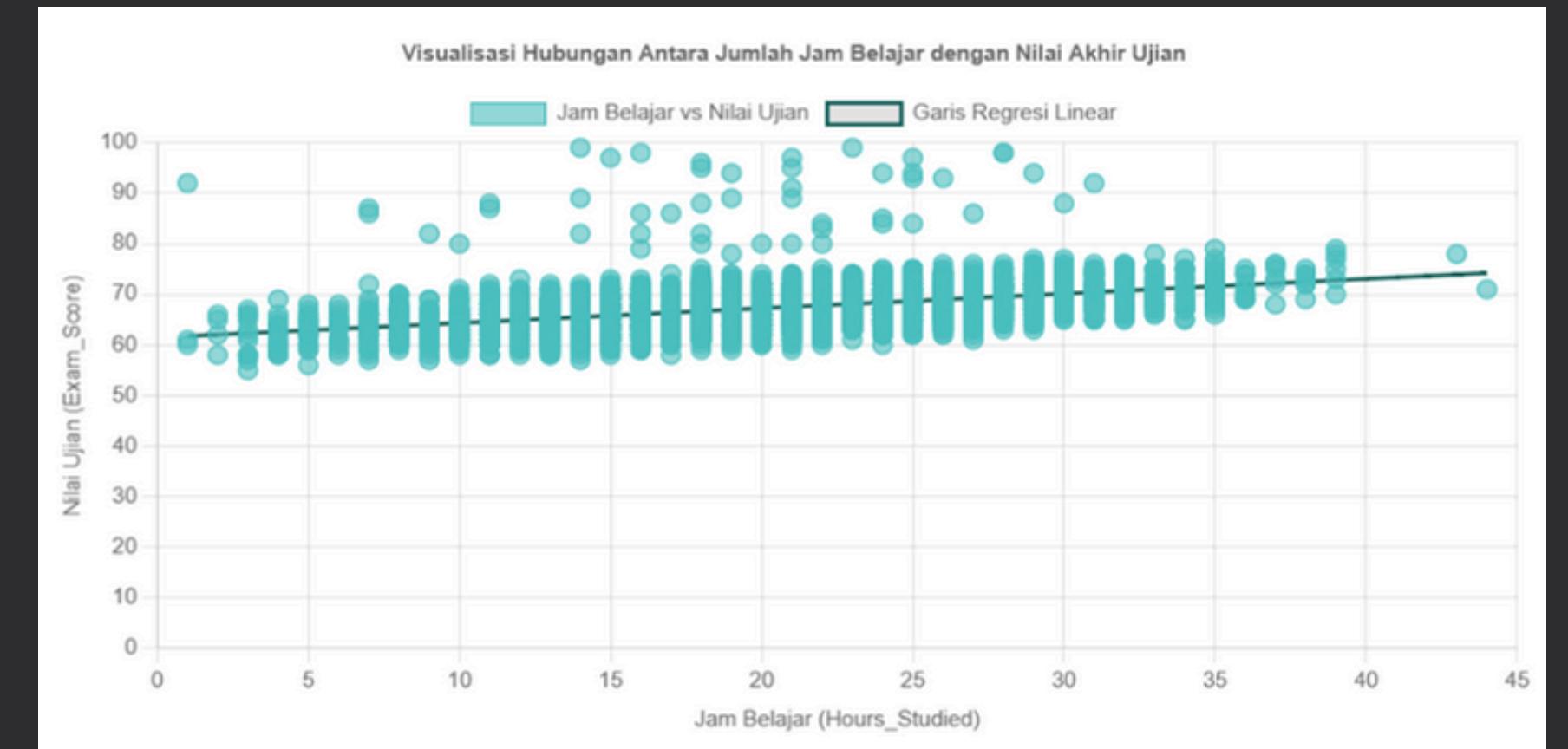
```
index.py  
df.duplicated().sum()  
Codelimage
```

```
df.duplicated().sum()
```

```
np.int64(0)
```

Can

SKETSA VISUALISASI



Insight Visualisasi

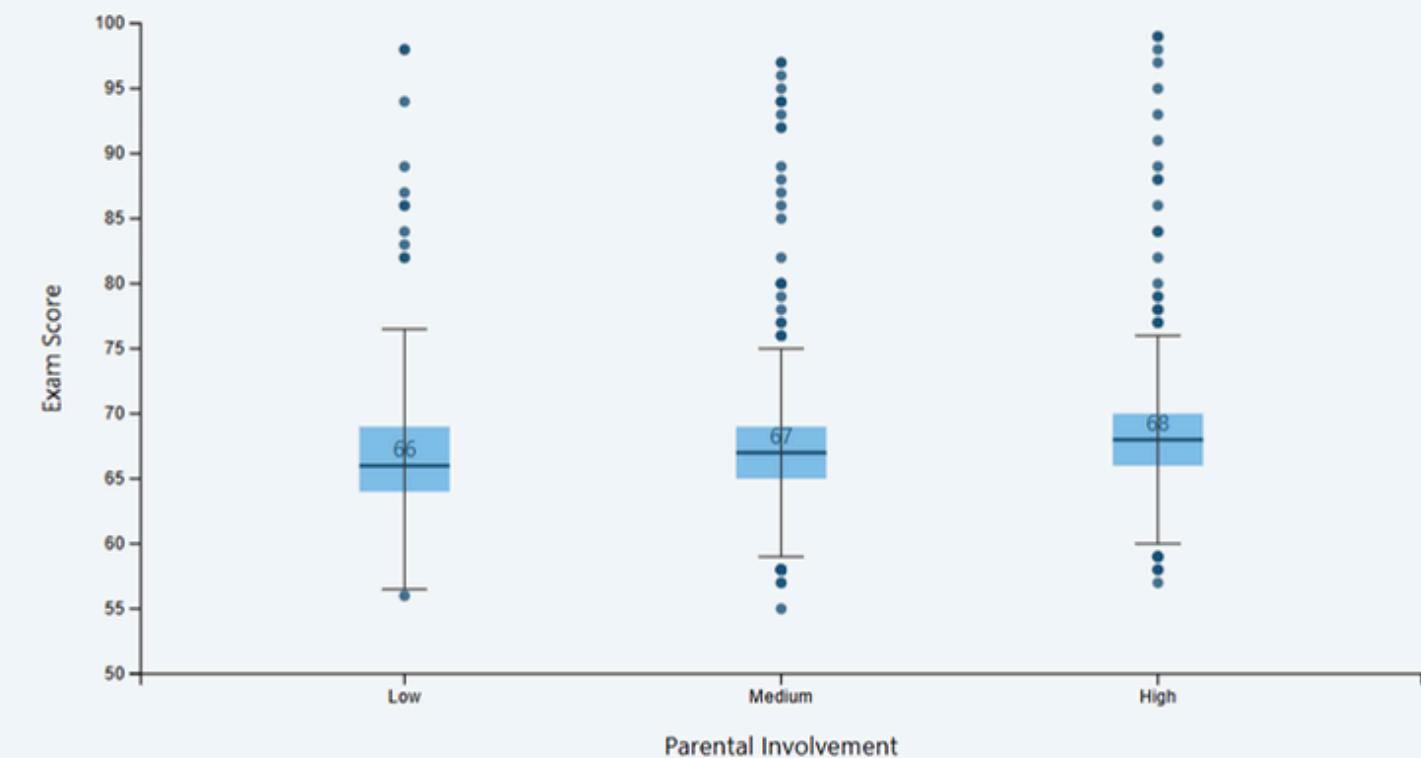
1. Terdapat korelasi positif namun lemah antara jam belajar dan nilai ujian, ditunjukkan oleh garis regresi yang miring sedikit ke atas, yang mengindikasikan bahwa peningkatan jam belajar hanya memberikan sedikit pengaruh terhadap kenaikan nilai ujian.
2. Sebagian besar nilai ujian berkisar antara 60-75 di seluruh rentang jam belajar (0-45 jam), menunjukkan bahwa durasi belajar bukanlah faktor dominan yang menentukan hasil ujian.
3. Ditemukan outlier nilai tinggi (di atas 85) pada berbagai tingkat jam belajar, bahkan pada jam belajar yang relatif rendah, yang mengindikasikan bahwa faktor-faktor lain mungkin lebih berpengaruh daripada jumlah waktu belajar yang dihabiskan.
4. Distribusi data menunjukkan variasi nilai yang lebar pada setiap interval jam belajar, mengkonfirmasi bahwa jumlah jam belajar saja tidak menjamin perolehan nilai yang lebih tinggi.

Insight Visualisasi

1. Terdapat korelasi positif antara keterlibatan orang tua dan nilai ujian, ditunjukkan dengan kenaikan nilai median yang konsisten dari kategori keterlibatan rendah (66), sedang (67), ke tinggi (68).
2. Meskipun ada peningkatan nilai ujian seiring meningkatnya keterlibatan orang tua, perbedaannya relatif kecil (hanya 1 poin di setiap level), menunjukkan bahwa pengaruhnya positif namun tidak dramatis.
3. Persebaran nilai (interquartile range) relatif mirip di ketiga kategori, namun kategori keterlibatan tinggi menunjukkan nilai kuartil atas yang lebih tinggi, mengindikasikan lebih banyak siswa dengan nilai di atas rata-rata.
4. Nilai ujian siswa outlier di atas 85 ditemukan di semua kategori keterlibatan orang tua, menunjukkan bahwa faktor lain selain keterlibatan orang tua juga berperan penting dalam pencapaian akademik siswa.

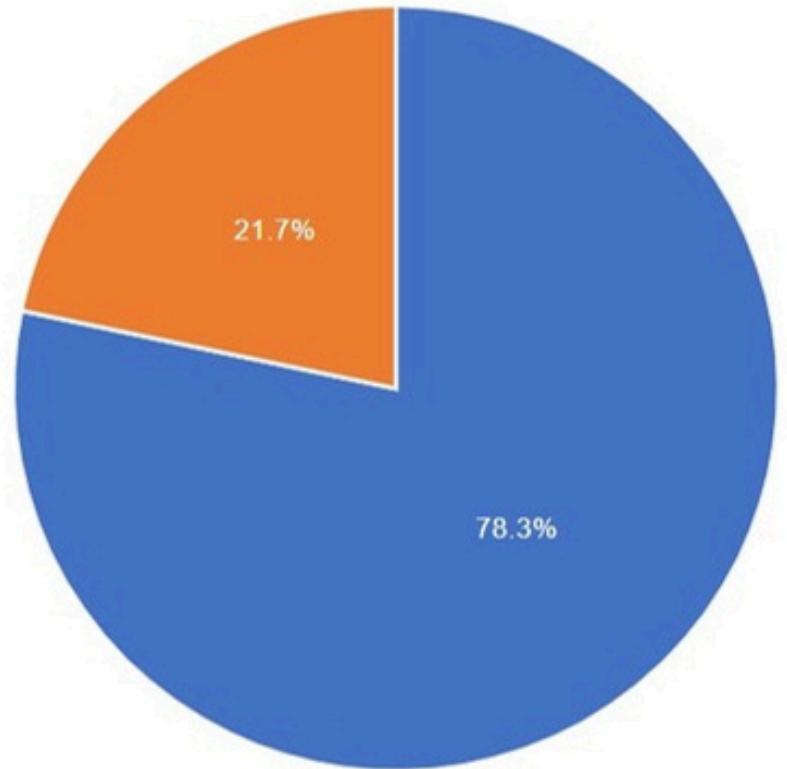
SKETSA VISUALISASI

Visualisasi Dampak Keterlibatan Orang Tua terhadap Performa Ujian Siswa



SKETSA VISUALISASI

Perbandingan Persentase Siswa dengan Nilai Tinggi Berdasarkan Jenis Sekolah



Insight Visualisasi

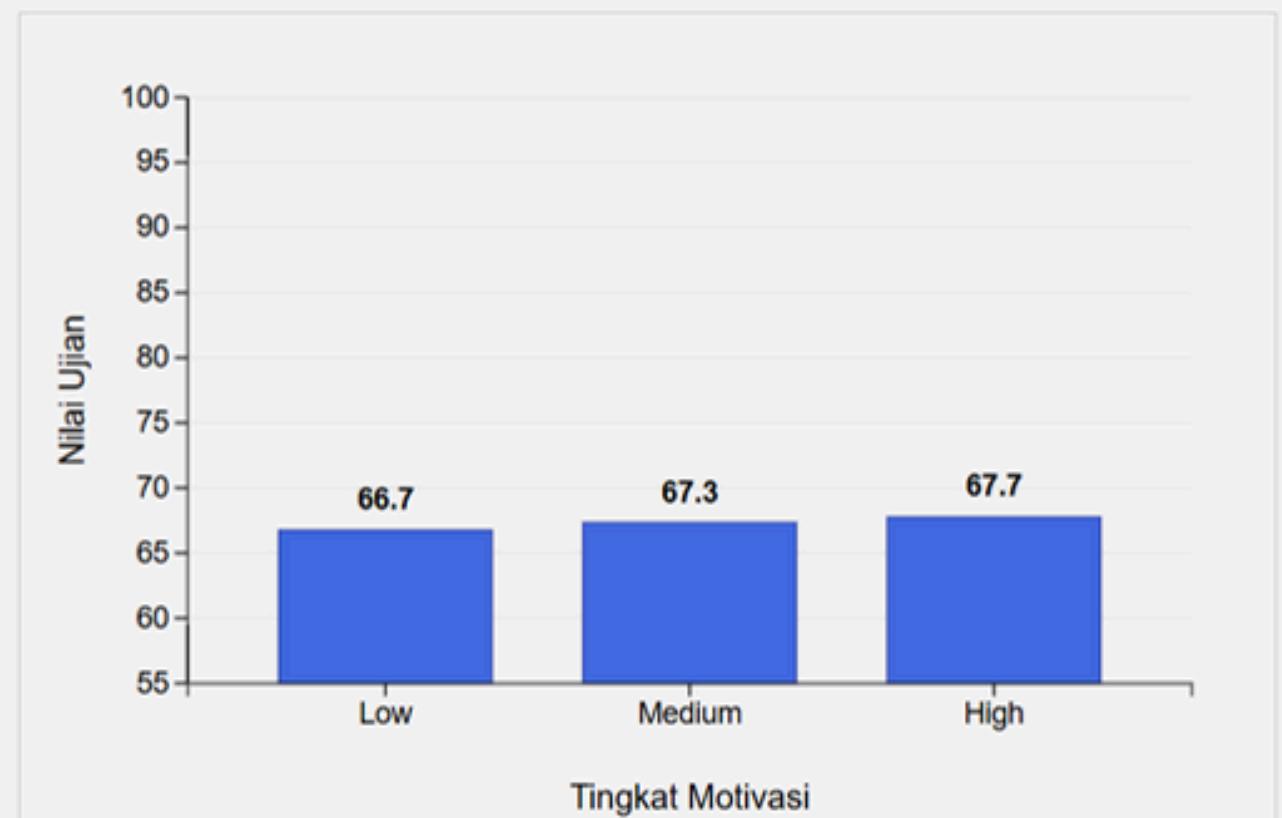
1. Sekolah publik mendominasi persentase siswa dengan nilai tinggi, mencapai 78,3% dari total, sementara sekolah swasta hanya menyumbang 21,7%.
2. Rasio perbandingan siswa berprestasi tinggi antara sekolah publik dan swasta adalah sekitar 3,6:1, menunjukkan kesenjangan yang cukup signifikan.
3. Kontribusi 10 siswa berprestasi dari sekolah swasta menunjukkan bahwa meskipun jumlah keseluruhan murid lebih sedikit dari sekolah publik, sekolah swasta tetap memiliki peran dalam menghasilkan siswa berprestasi tinggi.

Insight Visualisasi

1. Perbedaan hasil nilai antar tingkat motivasi tidak terlalu signifikan dan menandakan motivasi tidak terlalu berpengaruh terhadap nilai akhir siswa
2. Motivasi tinggi bukanlah loncatan drastis dalam capaian nilai, namun tetap membawa hasil tertinggi dibandingkan yang lainnya.
3. Faktor-faktor lain di luar motivasi kemungkinan memiliki pengaruh lebih besar terhadap nilai akhir, mengingat perbedaan nilai antar kelompok motivasi yang minimal.

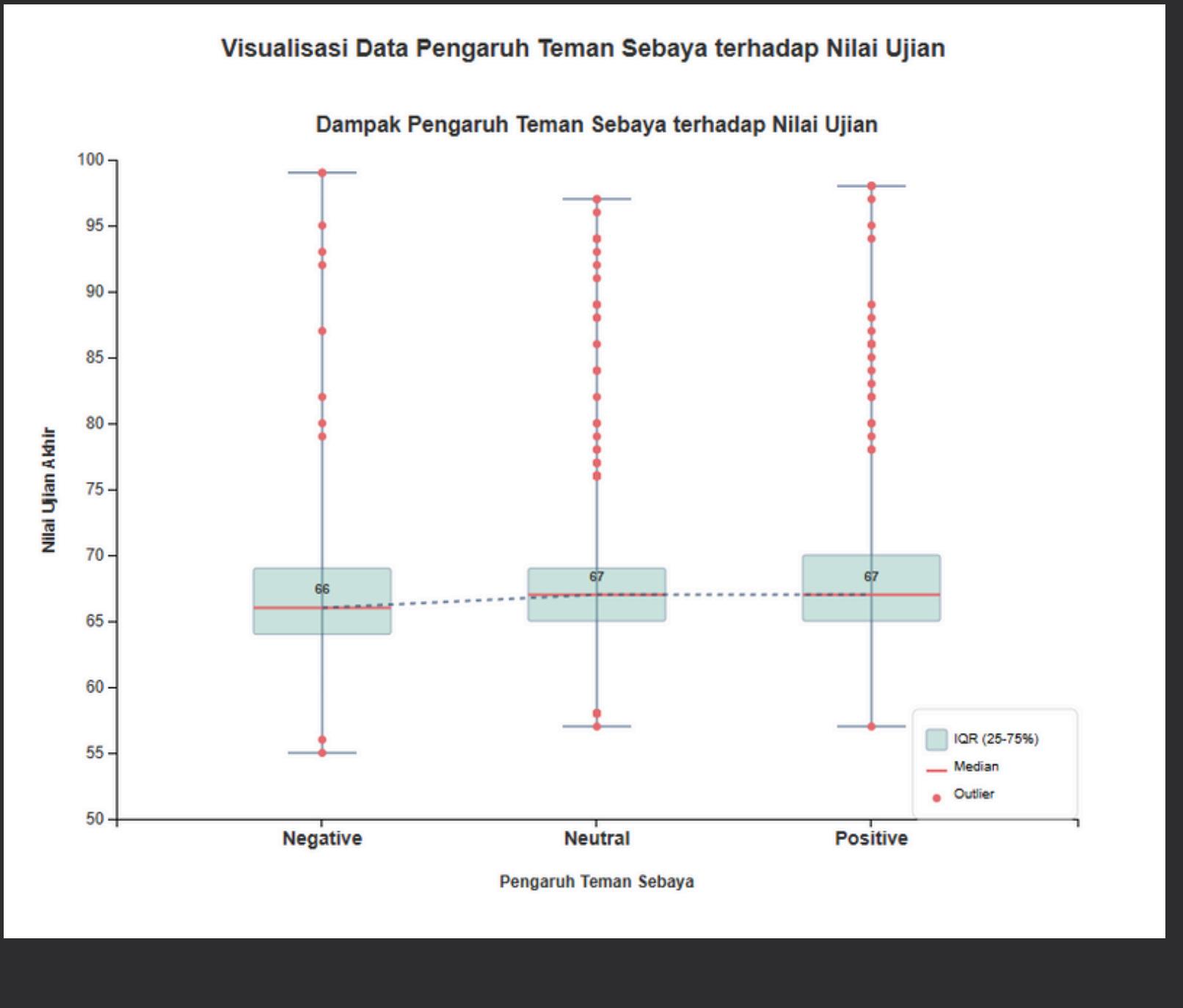
SKETSA VISUALISASI

Pengaruh Motivasi belajar siswa terhadap rata-rata nilai akhir



Insight Visualisasi

SKETSA VISUALISASI



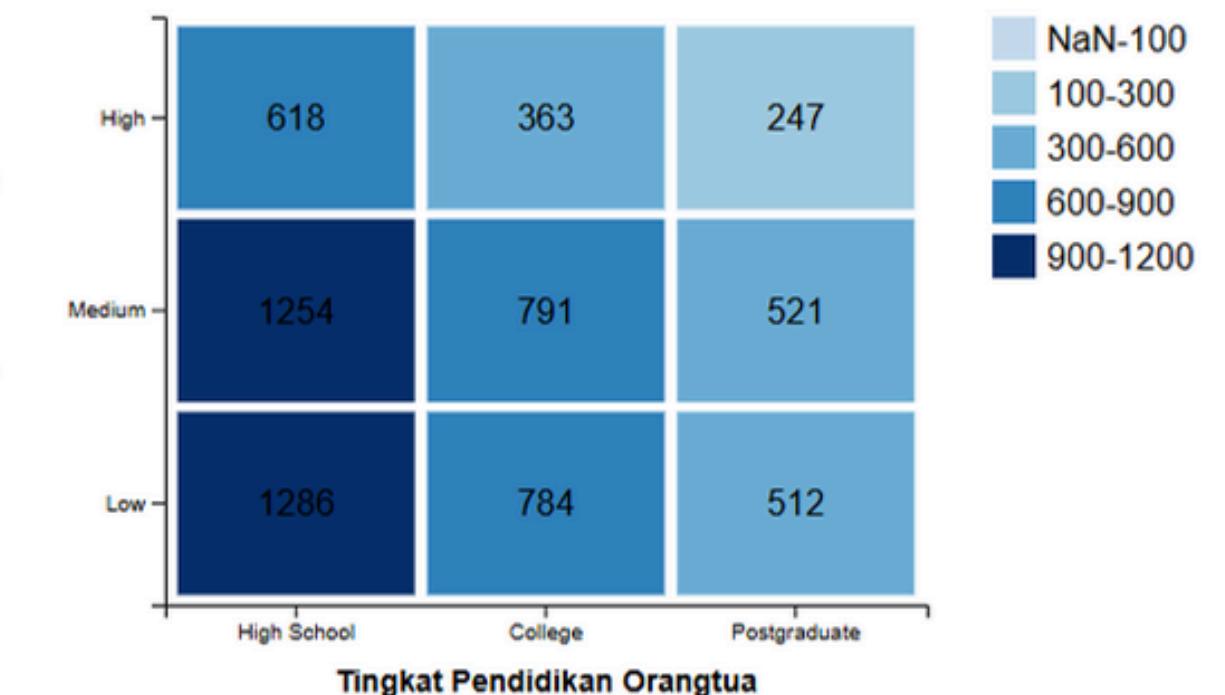
1. Tren peningkatan nilai median: Negative (66) → Neutral (67) → Positive (67)
2. Kelompok pengaruh Positive menunjukkan distribusi nilai lebih tinggi dan lebih terkonsentrasi
3. Siswa berprestasi tinggi tetap mencapai nilai outlier tertinggi di semua kelompok pengaruh
4. Nilai minimum lebih tinggi pada kelompok Neutral dan Positive dibanding Negative

Insight Visualisasi

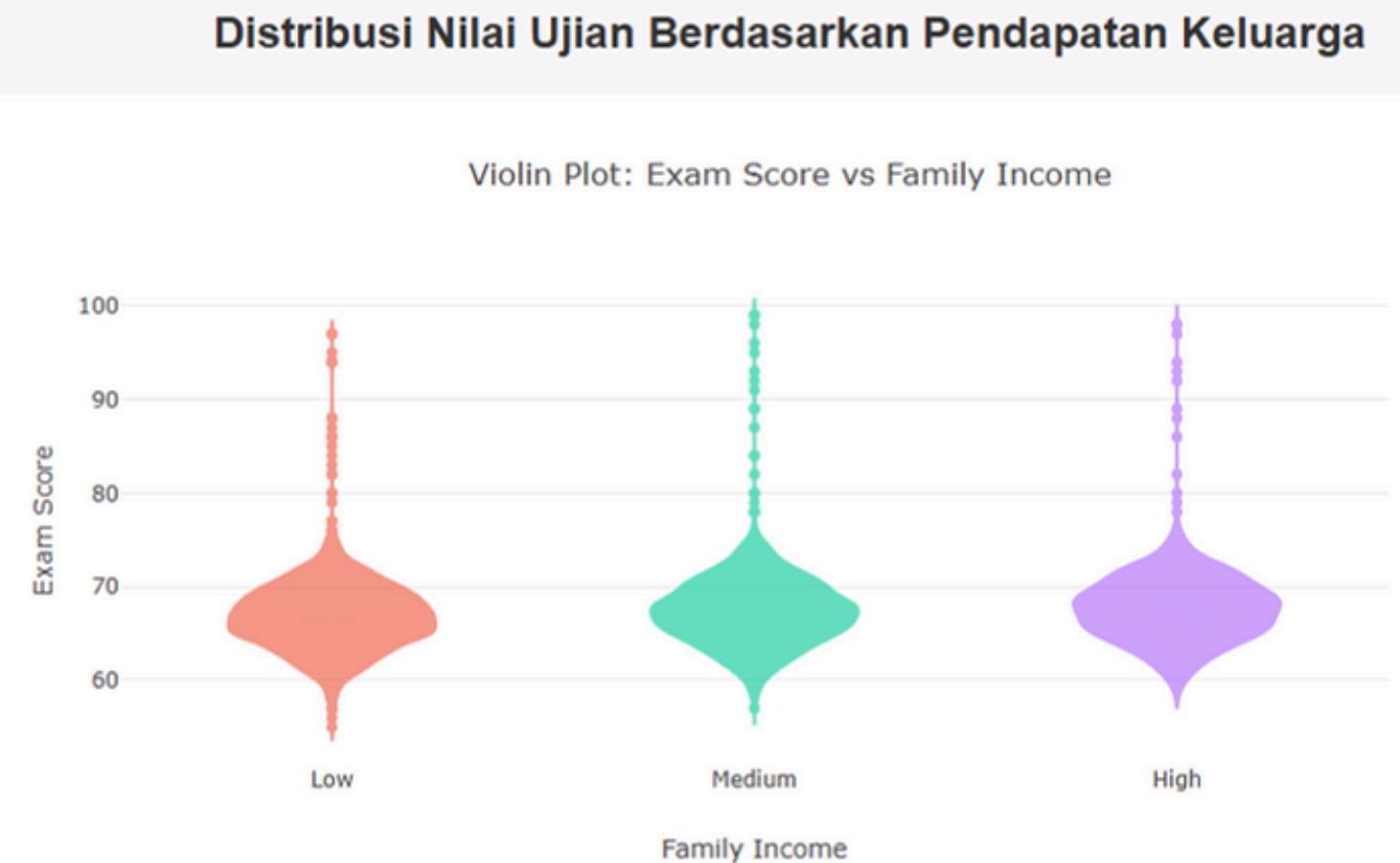
1. Terdapat korelasi antara pendidikan orang tua dan pendapatan keluarga, di mana orang tua dengan pendidikan setingkat SMA umumnya berada dalam kelompok pendapatan rendah
2. Pendidikan tinggi cenderung terkait dengan pendapatan yang lebih tinggi, namun bukan satu-satunya faktor, karena lulusan SMA masih mendominasi kelompok pendapatan tinggi.
3. Peningkatan kesejahteraan keluarga tidak hanya bergantung pada pendidikan formal, tetapi juga memerlukan dukungan ekonomi dan pelatihan keterampilan, terutama bagi keluarga dengan tingkat pendidikan rendah.

SKETSA VISUALISASI

Hubungan Tingkat Pendidikan Orangtua dengan Pendapatan Keluarga



SKETSA VISUALISASI



Insight Visualisasi

1. Tidak ada perbedaan signifikan dalam distribusi nilai ujian antar kelompok pendapatan keluarga (Low, Medium, High).
2. Kelompok berpendapatan Medium memiliki distribusi nilai yang lebih simetris dan konsisten, terutama di kisaran nilai 70-75.
3. Siswa dari keluarga berpendapatan menengah cenderung memiliki performa belajar yang lebih stabil dibanding kelompok lainnya.
4. Tingkat pendapatan keluarga bukan satu-satunya faktor yang memengaruhi performa belajar siswa, namun stabilitas performa lebih terlihat pada kelompok menengah.

Insight Visualisasi

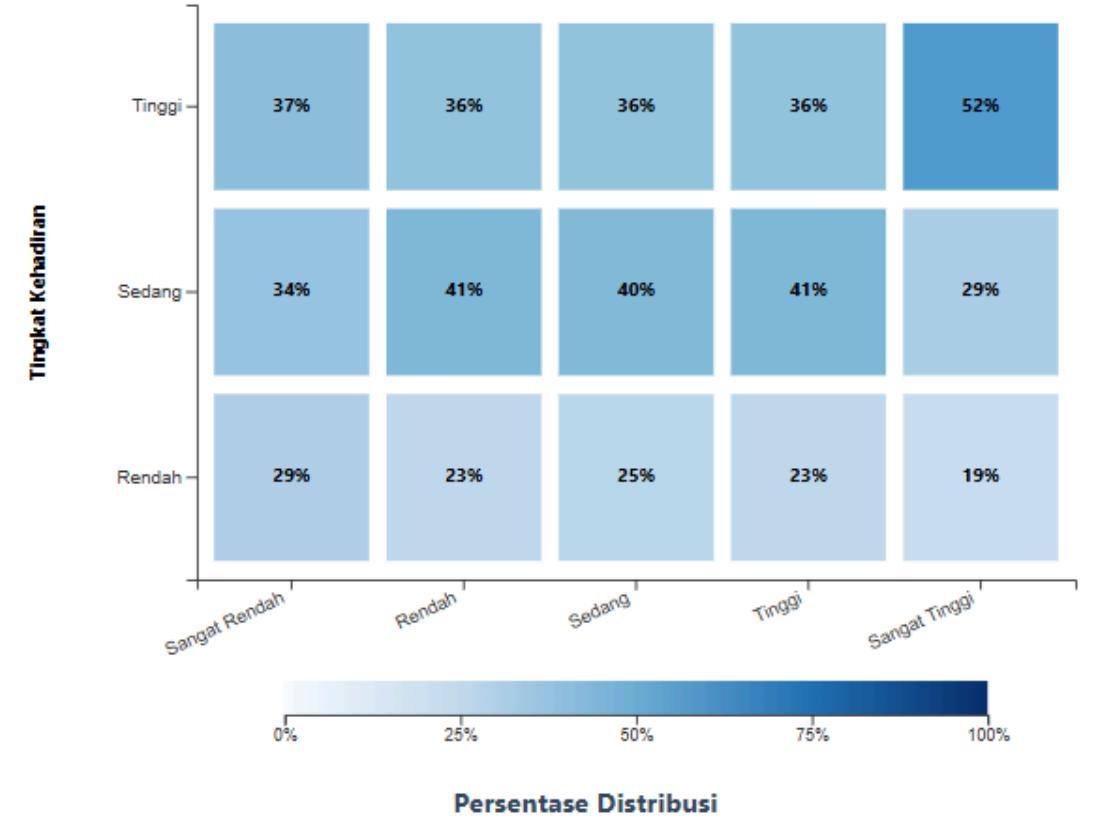
1. Jam belajar Sangat Tinggi (>35 jam) = kehadiran tinggi terbesar (52%)
2. Makin sedikit jam belajar = makin tinggi risiko kehadiran rendah
3. Kategori kehadiran Tinggi konsisten (36%) pada hampir semua kelompok jam belajar
4. Jam belajar Rendah/Tinggi sama-sama menghasilkan kehadiran Sedang (41%)
5. Jam belajar Sangat Rendah (<7 jam) = distribusi kehadiran paling merata
6. Semakin banyak jam belajar, semakin terpolarisasi pola kehadiran siswa

SKETSA VISUALISASI

Distribusi Jam Belajar dan Kehadiran Siswa

Visualisasi hubungan antara kategori jam belajar dan tingkat kehadiran

Distribusi Siswa berdasarkan Jam Belajar dan Kehadiran



Kategori Jam Belajar:

- Sangat Rendah: < 7 jam
- Rendah: 7 - 13 jam
- Sedang: 14 - 20 jam
- Tinggi: 21 - 35 jam
- Sangat Tinggi: > 35 jam

Kategori Kehadiran:

- Rendah: $< 70\%$
- Sedang: 70% - 85%
- Tinggi: $> 85\%$

KESIMPULAN

Analisis data menunjukkan bahwa prestasi dan aktivitas akademik siswa merupakan hasil interaksi kompleks berbagai aspek, bukan ditentukan oleh satu faktor dominan. Jam belajar, keterlibatan orang tua, motivasi, dan pengaruh teman sebaya menunjukkan korelasi positif terhadap yang lainnya. Akan tetapi, dari semua faktor yang ada, jam belajar relatif paling berpengaruh, dengan kategori sangat tinggi (>35 jam) menunjukkan tingkat aktivitas akademik tertinggi dari siswa yaitu kehadiran mereka yang mencapai 52%. Sementara pendapatan keluarga dan jenis sekolah tidak signifikan memengaruhi performa akademik maupun yang lainnya. Secara keseluruhan, keberhasilan akademik lebih ditentukan oleh kombinasi pembelajaran efektif dan karakteristik individual siswa.

THANK YOU!

