

Analisis Dan Visualisasi Data Menggunakan *Looker Studio* Pada Dataset *New York City Property Sales*

¹Indra Purnama, ^{2*}Yeni Setiani, ³Farhan Ari Nur Wibisono

¹²³ Prodi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi,
Universitas Gunadarma, Jakarta, Indonesia

¹indrapurnama@staff.gunadarma.ac.id, ^{2*}yeni_setiani@staff.gunadarma.ac.id,

³farhanarinw@gmail.com

ABSTRAK

Penulisan ilmiah ini berfokus pada analisis dan visualisasi dari dataset NYC Property Sales yang bersumber dari situs Kaggle. Penulisan ini berlatar belakang pada kemampuan untuk mengubah data kompleks dalam dataset menjadi informasi yang dapat digunakan untuk menyelesaikan masalah dan membuat keputusan yang tepat di masa depan. Analisis dan visualisasi ini digunakan dalam mengetahui permasalahan yang ada dalam dataset yaitu penurunan penjualan properti dan bertujuan untuk melakukan analisis dan visualisasi dengan hasil akhir berupa dashboard yang dibuat menggunakan Looker Studio. Metode analisis dan *Exploratory Data Analysis* (EDA) digunakan dalam penulisan ini dengan tahapan Data Collection, Data Preparation, EDA, dan Visualization. Dalam melakukan analisis dan visualisasi menghasilkan sebuah dashboard yang di dalamnya terdapat filter, scorecard, chart dan tabel. Dashboard yang telah dibuat dapat di akses melalui tautan berikut ini: <https://lookerstudio.google.com/reporting/a60afdae-b109-4043-99dc-1fa3f4b3df9f/page/bQ64D>. Dari pembahasan yang telah dilakukan, dashboard yang telah berhasil dibuat dapat memberikan informasi terkait tren waktu pada penjualan properti, rata-rata harga properti, penyebaran distribusi penjualan dan transaksi penjualan dalam suatu borough, kelas bangunan ataupun tahun pembuatan bangunan properti. Hal ini dapat memberikan informasi bagi para investor untuk melakukan investasi di waktu yang tepat serta memberikan wawasan yang berharga.

Kata Kunci: *Dashboard, Looker Studio, NYC Property Sales, Visualisasi Data*

PENDAHULUAN

Kebutuhan dalam memahami dan mengolah suatu data yang kompleks dalam suatu dataset menjadi informasi yang dapat digunakan untuk menyelesaikan suatu permasalahan dan pengambilan solusi atau keputusan di masa depan dengan tepat. Di era digital saat ini banyak data yang secara mentah dikumpulkan dan disimpan, namun bagaimana mengolah data tersebut sehingga dapat menjadi informasi yang bermanfaat terutama untuk pihak-pihak yang membutuhkan dalam mengambil keputusan yang tepat di masa depan dengan memprediksi sesuatu berdasarkan data yang telah diolah dan dianalisis menjadi informasi yang bermanfaat.

Analisa data dan memvisualisasikannya sangat penting bagi suatu bisnis dalam mengoptimalkan suatu bisnis dengan mengetahui suatu masalah yang ada dan diselesaikan dengan solusi berdasarkan data yang selanjutnya dianalisis dan menjadi informasi yang berguna. Dengan menganalisis data tersebut perusahaan akan dapat mengetahui tren dan pola yang ada pada data tersebut dan berguna dalam membantu perusahaan untuk memprediksi apa yang akan terjadi kedepannya. Menganalisis data juga dapat digunakan untuk memecahkan suatu permasalahan yang ada pada perusahaan dan meningkatkan penjualan serta keuntungan.

Oleh karena itu analisis dibuat dengan menggunakan metode analisis dan *Exploratory Data Analysis* (EDA) dalam mengolah data. Data diambil dari *website kaggle* yang merupakan sebuah komunitas online yang dibentuk oleh Anthony Goldbloom sebagai *Chief Executive Officer* (CEO) dan Ben Hamner sebagai *Chief Technology Officer* (CTO), pada tahun 2010 diakuisisi Google pada

tahun 2017. Kaggle diperuntukan bagi data scientists yang didalamnya menyediakan kompetisi *machine learning*, kumpulan data (Datasets), *notebooks*, akses ke akselerator pelatihan dan pendidikan. Melakukan data preparation dengan *spreadsheet* dan *bigquery*. Selanjutnya menggunakan *Looker Studio* dalam visualisasi data dengan menampilkannya dalam bentuk *dashboard*.

Ruang lingkup pada penelitian ini untuk menganalisis dataset *NYC Property Sales* yang diambil dari *website kaggle* yang berjumlah 84.549 data. Penelitian ini menggunakan metode data analisis dan *Exploratory Data Analysis* (EDA) serta pemvisualisasian data menggunakan *Looker Studio*. Rumusan permasalahan pada penelitian ini adalah; bagaimana proses analisis dari dataset *NYC Property Sales* dan pembuatan visualisasi menggunakan *Looker Studio*. Tujuan penulisan ilmiah ini adalah membuat visualisasi data berupa *dashboard* dengan menganalisis dataset untuk mengetahui permasalahan yang sedang dihadapi yaitu penurunan penjualan properti dari dataset *NYC Property Sales* dengan menggunakan *Looker Studio*.

TINJAUAN PUSTAKA

Looker Studio

Looker studio adalah alat *Business Intelligence* (BI) yang dirancang untuk mengumpulkan, menganalisis, dan memvisualisasikan data. *Looker studio* menawarkan berbagai fitur mulai dari analisis data yang mendalam, pembuatan laporan hingga menyajikan data secara *real-time* (Kurniawan, Yuliani & Pratama, 2023). *Google Looker Studio* adalah *platform* analisis berbasis *cloud* yang canggih untuk bisnis segala ukuran. Ini memberikan antarmuka yang intuitif dan mudah digunakan untuk membuat visualisasi dan laporan data yang kuat yang membantu organisasi mengidentifikasi tren, peluang, dan wawasan dalam data mereka. Dengan *Google Looker Studio*, bisnis dapat terhubung ke berbagai sumber data, termasuk *Google BigQuery*, *Salesforce*, dan *Oracle*. *Looker Studio* dirancang dengan mempertimbangkan pengguna akhir, memungkinkan pengguna menjelajahi data bisnis tanpa memerlukan pengkodean yang rumit. Dengan menggunakan fungsionalitas *drag-and-drop*, pengguna dapat membuat dasbor canggih dengan cepat dan mudah dengan menggabungkan berbagai jenis visual seperti bagan, peta, tabel dan lainnya. Setelah visual ini selesai, visual tersebut dapat dengan mudah dibagikan dengan anggota tim atau pelanggan lain melalui email atau *platform online* lainnya.

Beberapa penelitian yang sudah pernah melakukan *tools looker studio* antara lain; (Darmanto et al., 2023), menjelaskan bahwa menggunakan *looker studio* kapasitas penyimpanan dan jumlah file menjadi lebih efektif dan rapih dikarenakan untuk menambah data tidak perlu mengupdate filenya, tetapi cukup dengan menambah baris datanya saja. Akses terhadap data produksi menjadi mudah dan informatif sehingga dapat mempercepat pengambilan keputusan, tindakan ataupun menentukan kebijakan untuk meningkatkan kinerja perusahaan. Selain itu, variasi sumber data *Looker studio* memungkinkan analisis data yang dihasilkannya lebih akurat dan mendalam. Proses memodelkan data *Looker* menggunakan *LookML* (*Looker Modeling Language*), yang memungkinkan pengguna mendefinisikan metrik dan dimensi dengan cara yang lebih terstruktur (Prasetyo & Handayani, 2021). Penelitian yang dilakukan oleh (Fitri Ariani et al., 2024) implementasi ke dalam *Google Sheets* dan *Looker Studio* untuk pembuatan *dashboard* interaktif. Hasil penelitian menunjukkan bahwa *dashboard* interaktif yang dikembangkan mampu menampilkan data yang tepat dan *realtime*, sementara *dashboard* ini memberikan wawasan mendalam mengenai objek yang menjadi fokus penelitian dan memudahkan pengguna dalam pengambilan keputusan berbasis data. Fitur interaktivitas seperti filter dan *drill-down* memungkinkan eksplorasi data yang lebih rinci, memberikan nilai tambah bagi pengambil kebijakan dan pemangku kepentingan lainnya.

Visualisasi Data

Cara penyampaian data yang menarik dan mudah dipahami dengan grafik adalah visualisasi data (Irmayani, 2021). Visualisasi data membantu menemukan pola, tren, dan anomali yang tidak terlihat dalam tabel data tradisional (Sartono & Hidayatuloh, 2012). Pola antar variabel dalam data ditampilkan, yang mempercepat proses pengambilan keputusan bagi pengguna. Beberapa jenis visualisasi data yang paling umum adalah grafik, diagram, dan grafik. Grafik menunjukkan

hubungan antara dua variabel, dan diagram menunjukkan gambaran umum proses. Selain itu, diagram dapat digunakan untuk visualisasi data dan sesuai dengan informasi yang ingin ditampilkan. Gambar adalah visualisasi yang menunjukkan hubungan yang lebih ringkas. Ketiganya dapat disesuaikan dalam hal fungsi dan tampilan.

Dataset

Dataset adalah suatu database didalam memori (*in-memory*). Dataset memiliki semua karakteristik, fitur dan fungsi dari database biasa. Dataset dapat memiliki banyak tabel dan tabel dapat memiliki hubungan (*relationship*). Tabel-tabel pada suatu dataset dapat memiliki foreign key dan integritas referensial.

Dataset adalah objek yang merepresentasikan data dan relasinya di memory. Strukturnya mirip dengan data yang ada di database. Dataset berisi koleksi dari data tabel dan data (Yahya & Mahpuz, 2019).

METODE PENELITIAN

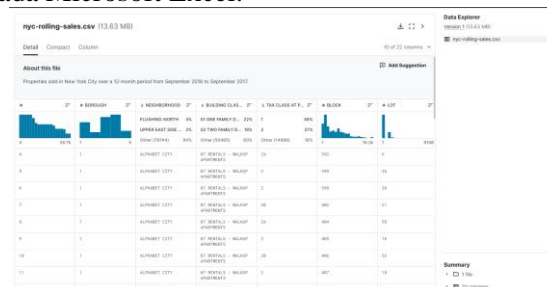
Pada metode penelitian ini dilakukan tahapan yang dapat dijelaskan sebagai berikut:

1. Pengumpulan Data (*Data Collection*)
Pada tahapan ini, dilakukan sebuah pengumpulan data yang relevan dengan sebuah permasalahan yang ingin diteliti oleh penulis, melalui tahapan ini diperoleh sebuah *dataset* yang didalam nya terdapat data mengenai *New York City Property Sales* yang bersumber dari situs *kaggle*.
2. Pra-pemrosesan Data (*Data Preprocessing*)
Pada tahap ini, data mentah akan disiapkan untuk menjadi data yang berkualitas. Pada tahap ini dilakukan format data, penanganan duplikasi data, penanganan missing value, penanganan whitespace, penanganan data tidak penting, dan penanganan outliers.
3. Exploratory Data Analysis (EDA)
Pada tahapan ini dilakukan eksplorasi terhadap dataset yang ingin dianalisis dengan menggunakan *BigQuery*. Metode EDA mengacu pada investigasi awal yang kritis terhadap data untuk menemui pola, mendeteksi anomali, menguji hipotesis, dan memverifikasi asumsi dengan bantuan statistik deskriptif dan representasi grafis.
4. Visualisasi Data (*Visualization*)
Pada tahapan terakhir adalah melakukan *visualisasi* data, data yang telah melalui proses yang dilakukan pada *Data Preprocessing* dan EDA dapat divisualisasikan dengan dashboard menggunakan *Looker Studio*.

HASIL DAN PEMBAHASAN

1. Pengumpulan Data (*Data Collection*)

Menurut Yeni (2023), Pengumpulan data adalah suatu proses sistematis dalam mengumpulkan hasil pengamatan atau pengukuran, yang memungkinkan individu dalam memperoleh pemahaman langsung mengenai masalah penelitian. Pada tahap data collection dilakukan pengambilan data yang bersumber dari situs *kaggle*. Untuk mengolah data ini dilakukan *download* file dengan format CSV yang berisi dataset NYC *Property Sales* yang selanjutnya akan diolah pada Microsoft Excel.



Gambar 1. Dataet-nyc-rolling-sales.csv

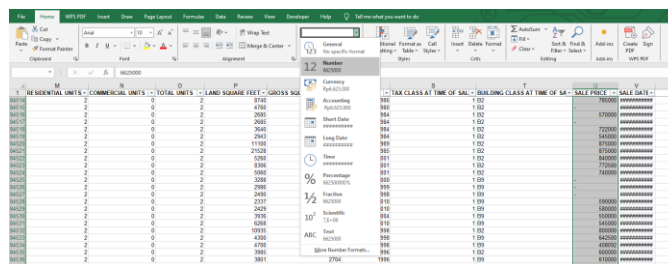
2. Pra-pemrosesan Data (*Data Preprocessing*)

Menurut Fauzan (2021), Persiapan data adalah langkah awal yang penting dalam menyiapkan data mentah menjadi data yang berkualitas untuk digunakan pada tahap selanjutnya.

Pada tahapan data preprocessing dilakukan pengelolaan dan pembersihan data yang telah dimasukkan pada Microsoft Excel seperti format data, penanganan duplikasi data, penanganan nilai yang hilang (*Missing value*), penanganan spasi berlebih (*Whitespace*), penanganan data tidak penting, dan penanganan *outliers*. Tahapan-tahapan yang dilakukan dilakukan agar pada saat menganalisis dan visualisasi data tidak terjadi masalah dan hambatan serta memberi kemudahan dalam melakukannya. Pengelolaan dan pembersihan data dilakukan dengan menggunakan alat yang tersedia pada bawaan Microsoft Excel.

2.1 Format Data

Pada langkah format data dilakukan penyeragaman atau perbaikan tipe data yang ada pada tiap kolom dengan mengubahnya kedalam tipe data yang sesuai dengan kebutuhan. Pengubahan data yang berbentuk kode berupa angka juga diubah menjadi bentuk angka.

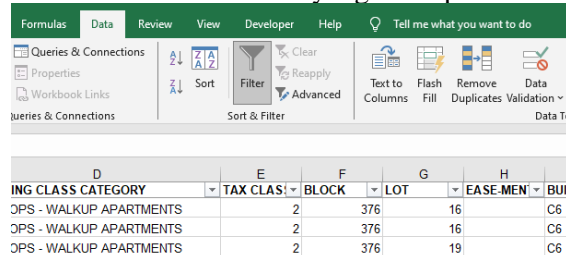


Gambar 2. Perbaikan Format Pada Kolom *Sale_Price*

Pada kolom *Sale_Price* yang akan menjadi acuan utama dalam melakukan analisis dan visualisasi data dilakukan perbaikan format. Untuk melakukan pengecekan dan perbaikan format data pada kolom *Sale_Price* yang telah di blok sepenuhnya dengan menggunakan alat yang ada pada bagian Home, Number. Terdapat kotak persegi panjang yang dapat diklik lalu memilih *format number*. Melakukan pengecekan dan perbaikan format data pada kolom *Sale_Price* yang telah di blok sepenuhnya dengan menggunakan alat yang ada pada bagian Home, Number. Terdapat kotak persegi panjang yang dapat diklik lalu memilih format *Short Date*. Melakukan perubahan kode pada kolom *borough* menjadi nama *borough*. Penjelasan terkait kode dan nama borough ada pada deskripsi sumber data. Rincian terkait kode dan nama borough sebagai berikut : angka 1 untuk Manhattan, angka 2 untuk Bronx, angka 3 untuk Brooklyn, angka 4 untuk Queens, angka 5 untuk Staten Island. Hal ini dilakukan untuk memudahkan saat dilakukan analisis lebih lanjut.

2.2 Penanganan Duplikasi Data

Pada langkah selanjutnya melakukan penanganan pada duplikasi data. Hal ini dilakukan agar hasil dari analisis dan visualisasi bersih dari data yang sama pada baris yang berbeda.

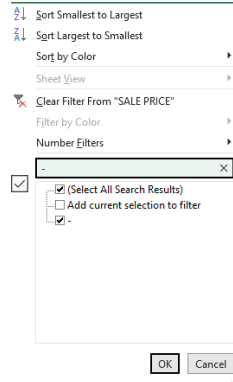


Gambar 3. Pengecekan Duplikasi

Melakukan pengecekan duplikasi melalui Microsoft Excel dengan menggunakan alat *Remove Duplicates* yang ada pada bagian Data. Setelah dilakukan *Remove Duplicates* diketahui bahwa terdapat 385 duplikasi data dan data tersebut telah dihapus.

2.3 Penanganan Nilai Kosong (*Missing Value*)

Pada langkah selanjutnya dilakukan penanganan pada nilai yang kosong (*Missing Value*). Penanganan nilai kosong ini dilakukan agar tidak terjadi permasalahan dalam melakukan analisis dan visualisasi.

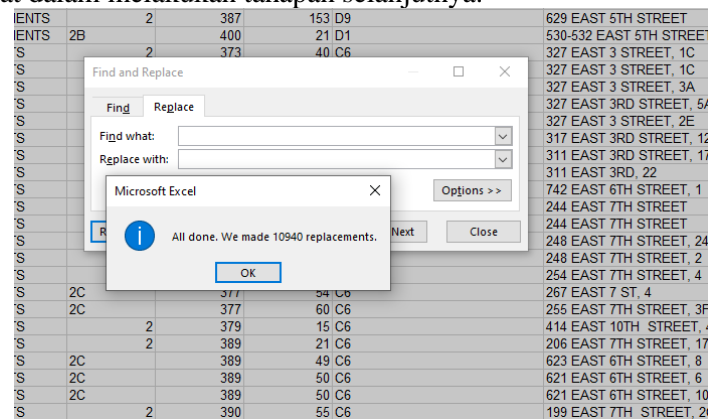


Gambar 4. Pengecekan *Missing Value* Pada Kolom *Sale_Price*

Untuk melakukan pengecekan *Missing value* pada kolom *sale_price* dengan memastikan bagaimana bentuk *missing value* yang ada pada kolom *sale_price*. *Missing value* yang ada pada kolom *sale_price* ini berbentuk dalam strip (-). Pengecekan dilakukan dengan menggunakan alat *filter* yang ada pada bagian *Data, Sort & Filter*. Setelah dicek terdapat 14.561 dari 84.547 data yang nilainya hilang dalam kolom *sale_price*. Data yang hilang tersebut dihapus dengan blok semua data yang telah difilter lalu klik kanan dan *delete row*.

2.4 Penanganan Spasi Berlebih (*Whitespace*)

Pada langkah Selanjutnya melakukan penghapusan white space atau spasi berlebih pada data. Penghapusan ini dilakukan agar saat melakukan analisis dan visualisasi tidak terjadi masalah yang akan menghambat dalam melakukan tahapan selanjutnya.

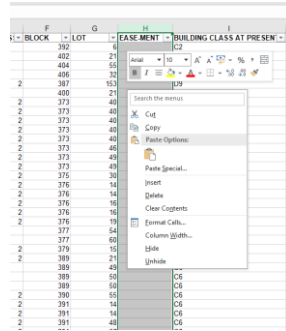


Gambar 5. Penghapusan *Whitespace*

Untuk melakukan penghapusan spasi berlebih pada data dilakukan dengan menggunakan cara manual yaitu dengan menggunakan alat *Find and Replace* yang ada pada bagian *Home, Editing*. Caranya dengan melakukan dua kali spasi pada bagian *Find what* dan dua kali spasi pada bagian *Replace with* lalu klik *Replace all* sampai semua data yang ada spasi berlebih tidak ditemukan. Dalam melakukan penghapusan *white space* pada data ditemukan 10.940 data yang terdapat spasi berlebih pada percobaan pertama.

2.5 Penanganan Data Tidak Penting

Pada langkah selanjutnya melakukan penanganan data yang tidak penting atau tidak digunakan dalam analisis dan visualisasi nantinya. Data yang tidak penting mencakup data kosong pada suatu kolom ataupun data yang memang tidak dibutuhkan dalam analisis dan visualisasi.



Gambar 6. Penghapusan Kolom *Easement*

Melakukan penghapusan kolom *Easement* dengan alasan kolom tersebut tidak memiliki nilai sama sekali atau kosong. Penghapusan dilakukan dengan memblok kolom *Easement* lalu klik kanan dan *Delete*. Melakukan penghapusan kolom yang tidak ada namanya atau dikenal dengan *unknown name*. Penghapusan kolom dilakukan dengan alasan kolom tersebut dinilai hanya memiliki data yang berisi penomoran dan tidak memiliki nilai penting didalamnya. Penghapusan dilakukan dengan memblok kolom *unknown name* lalu klik kanan dan *Delete*. Melakukan penghapusan nilai 0 sampai dengan 20 pada data *Sale_price* yang berjumlah 10.891 data dengan alasan nilai tersebut merupakan nilai yang terjadi akibat pemindahan kepemilikan properti seperti pemindahan kepemilikan properti dari orang tua kepada anaknya. Hal ini dijelaskan dalam deskripsi yang ada pada tempat pengambilan data yaitu Kaggle. Terdapat 4205 data pada kolom *year_built* yang akan mengganggu proses analisis dan visualisasi nantinya. Untuk itu menghapusnya merupakan alasan yang tepat agar analisis dan visualisasi dapat berjalan dengan baik.

2.6 Penanganan *Outlier*

Pada langkah selanjutnya melakukan pengecekan *outlier* dan menanganinya. Pengecekan *outlier* dilakukan pada data kolom *sale_price* dengan menggunakan IQR (*Interquartile Range*). Setelah dilakukan pengecekan *outliers* maka *outliers* tersebut akan dihapus dari data yang akan dianalisis dan divisualisasikan.

Tabel 1. Pengecekan *Outlier*

Mengecek <i>Outlier</i>	
Q1	382425,75
Q3	1050000
IQR	667574,25
Batas Bawah	-618935,625
Batas Atas	2051361,375

Untuk melakukan pengecekan *outlier* pada data *Sale_price* dengan menggunakan IQR pada data yang telah dibersihkan sebelumnya. Pada tabel 1 hasil dari nilai Q1 atau kuartil 1 didapat dengan menggunakan rumus yang ada pada Microsoft Excel yaitu “=QUARTILE.EXC(S2:S58712;1)”. Penjelasan dalam rumus tersebut adalah menghitung kuartil 1 dengan mengetikkan 1 diakhir dan memblok data yang akan dihitung yang nilainya seperti pada S2:S58712. Hasil nilai dari Q3 atau kuartil 3 didapat dengan menggunakan rumus yang sama pada kuartil 1 namun bedanya hanya angka pada akhir rumus yang diketikkan angka 3 seperti berikut “=QUARTILE.EXC(S2:S58712;3)”. Hasil IQR didapat dari pengurangan antara hasil Q3 dengan hasil Q2. Hasil dari bawah bawah didapat dengan menggunakan rumus $Q1 - (1,5 * IQR)$ dan batas atas didapat dengan menggunakan rumus $Q3 + (1,5 * IQR)$.

SALE PRICE	SALE DATE	Outlier
6625000	19/07/2017	Outlier
3936272	23/09/2016	Outlier
8000000	17/11/2016	Outlier
3192840	23/09/2016	Outlier
16232000	07/11/2016	Outlier
10350000	17/10/2016	Outlier
499000	10/03/2017	Bukan Outlier
529500	09/06/2017	Bukan Outlier
423000	14/07/2017	Bukan Outlier
501000	16/03/2017	Bukan Outlier
450000	01/09/2016	Bukan Outlier
510000	17/08/2017	Bukan Outlier
350000	30/08/2017	Bukan Outlier
11900000	21/06/2017	Outlier
11900000	21/06/2017	Outlier
350000	30/01/2017	Bukan Outlier
404000	07/07/2017	Bukan Outlier

Gambar 7. Pembuatan Kolom Outlier

Membuat kolom *outlier* dengan tujuan agar dapat memfilter data mana saja yang mengandung *outlier*. Isi dari kolom *outlier* merupakan pengecekan terhadap data yang outlier dengan menggunakan rumus IF yang telah disediakan oleh Microsoft Excel. Rumus IF yang digunakan seperti ini:

"=IF(S2<\$X\$8;"Outlier";IF(S2>\$X\$9;"Outlier";"Bukan Outlier"))"

Penjelasan dari rumus IF tersebut adalah jika kolom yang dipilih memiliki nilai yang kurang dari batas bawah yang dimana nilai dari batas bawah ada pada cell X8 maka akan mengeluarkan nilai *Outlier* namun apabila tidak maka akan dicek lagi jika kolom yang dipilih nilainya lebih dari batas atas yang dimana nilai dari batas atas ini ada di cell X9 maka akan mengeluarkan nilai *outlier*, namun apabila tidak maka akan mengeluarkan nilai *Bukan Outlier*. Penghapusan data yang mengandung *outlier* dilakukan dengan memfilter data yang bernilai *outlier* pada kolom *outlier*. Lalu data yang terfilter akan dihapus keseluruhannya. Terdapat 5.870 data *outlier* yang dihapus dan data yang tersisa setelahnya berjumlah 48.636 data. Menghapus kolom *Outlier* dengan memblok kolom tersebut secara keseluruhan lalu klik kanan dan *delete*. Kolom *outlier* dihapus karena data yang mengandung *outlier* sudah dihapus yang menjadikan kolom *outlier* tidak dibutuhkan lagi. *Dataset* yang dihasilkan dalam tahap ini akan digunakan untuk tahapan selanjutnya.

3. Exploratory Data Analysis (EDA)

Menurut Feliks (2021), *eksploratory data analysis* (EDA) adalah tahapan fundamental dalam metodologi sains data yang bertujuan untuk memahami dan mengeksplorasi karakteristik dari suatu dataset. Pada tahapan *exploratory data analysis* dilakukan eksplorasi data secara mendalam dengan melakukan statistika deskriptif yang digunakan sebagai pemahaman awal terhadap data yang akan dianalisis dan divisualisasikan, mengimport dataset yang dihasilkan pada tahap data *preprocessing* ke dalam *Google Sheets* dan dihubungkan ke dalam *BigQuery*, dan melakukan eksplorasi atau pencarian pola serta tren yang ada dalam data NYC *Property Sales* dengan *queri* yang memakai bahasa SQL di dalam *BigQuery*.

3.1 Statistika Deskriptif

Melakukan statistika deskriptif terhadap data pada kolom *sale_price* yang digunakan sebagai pemahaman awal terhadap data yang digunakan untuk analisis dan visualisasi. Dalam melakukannya dilakukan penghitungan terhadap jumlah data (*count*), nilai minimum, nilai maksimum, rata-rata (*mean*), nilai tengah (*median*), nilai paling banyak (*mode*), dan kecondongan (*Skewness*).

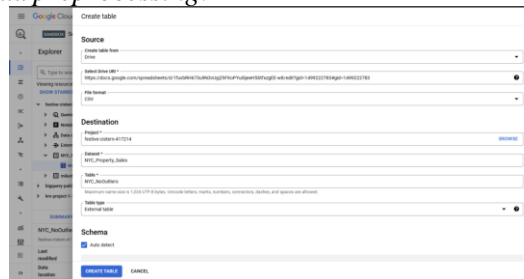
Tabel 2. Statistika Deskriptif

Sale Price	
Statistika Deskriptif	Nilai
Count	48636
Minimum	100
Maximum	2050000
Mean	660093
Median	570000
Mode	450000
Skewness	1,028881445

Untuk melakukan statistika deskriptif dengan data *sale_price* dilakukan penghitungan dalam mencari jumlah data (*count*) dengan rumus “=COUNT()”, nilai minimum dengan rumus “=MIN()”, nilai maksimum dengan rumus “=MAX()”, rata-rata (*mean*) dengan rumus “=AVERAGE()”, nilai tengah (median) dengan rumus “=MEDIAN()”, nilai paling banyak (mode) dengan rumus “=MODE()”, dan kecondongan (*Skewness*) dengan rumus “=SKEW()”. Semua rumus Microsoft Excel yang disebutkan dilakukan dengan menuliskan rumus lalu memilih data yang ingin dimasukkan kedalam rumus dengan memblok seluruh datanya. Hasil dari penghitungan tersebut dapat dilihat pada tabel 2.

3.2 Menghubungkan Sumber Data

Melakukan pengimporan dataset yang telah dihasilkan pada tahap *preprocessing* ke dalam *Google Sheet* yang selanjutnya akan dilakukan penghubungan ke dalam *BigQuery*. Menghubungkan sumber data dilakukan agar mempermudah dalam penyambungan ke *Looker Studio* dan mengubah data dimasa yang akan datang. Dari sumber data yang dihasilkan pada tahap data *preprocessing* dilakukan pengimporan ke dalam *Google Sheet*. Pengimporan data dilakukan dengan cara membuka *Google Sheet* lalu klik File, Import, *upload* dan pilih data yang dihasilkan pada tahap data *preprocessing*.

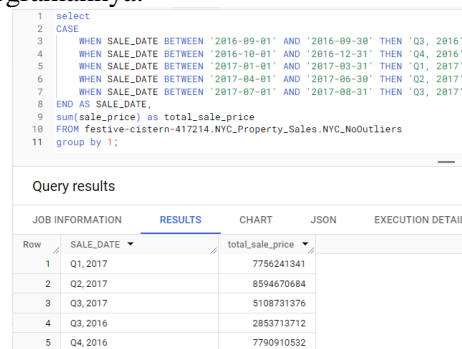


Gambar 8. Penyambungan *File Google Sheets* Ke Dalam *BigQuery*

Penyambungan dilakukan dengan membuat dataset terlebih dahulu dengan nama *NYC_Property_Sales* lalu membuat tabel dengan sumber data yang berasal dari *drive* disertai dengan linknya. Mengisi pilihan format menjadi CSV. Tabel yang dibuat bernama *NYC_NoOutliers*. Selanjutnya dilakukan eksplorasi data menggunakan *Query* dalam bahasa SQL.

3.1 Pencarian Pola Dan Tren

Pada langkah selanjutnya melakukan eksplorasi data dengan pencarian pola dan tren yang ada dalam dataset *NYC Property Sales*. Pencarian dilakukan dengan menggunakan *BigQuery* dan SQL sebagai bahasa pemrogramannya.



Gambar 9. Total Penjualan Per Kuartal

Melakukan pencarian total penjualan per kuartal. Dalam *query* pada gambar 9 menggunakan perintah *select* yang menampilkan data *sale_date* dengan *case* sebagai pemilihan rentang data yang diubah kedalam kuartal yang diberi nama *sale_date*. Menampilkan total penjualan pada data *sale_price* dengan perintah *sum* yang diberi nama *total_sale_price*. Hasil dari menjalankan *query* tersebut adalah menampilkan total penjualan per kuartal.

Melakukan pencarian total transaksi secara keseluruhan dari tiap *borough*. Dalam *query* pada gambar diatas menggunakan perintah *select* yang menampilkan data *borough*, data *sale_price*

dengan penghitungan *count* yang diberi nama *total_sales*. Hasil dari menjalankan *query* tersebut yaitu menampilkan data total transaksi pada tiap *borough*.

Melakukan pencarian total transaksi berdasarkan data *year_built* atau berdasarkan tahun pembuatan properti. Dalam *query* pada gambar diatas menggunakan *select* yang akan menampilkan data *year_built* dan banyaknya jumlah *year_built* dengan penghitungan *count* yang diberi nama *total_unit_sales*. Lalu data tersebut akan ditampilkan dengan urutan terbesar ke terkecil. Menggunakan limit 10 karena hanya akan mencari 10 tahun pembuatan yang memiliki total transaksi terbesar. Hasil dari menjalankan *query* tersebut yaitu menampilkan 10 data terbesar dari *year_built* atau tahun pembuatan properti beserta total transaksinya yang diurutkan dari yang terbesar.

Melakukan pencarian rata-rata penjualan berdasarkan borough dengan rentang waktu yang digunakan yaitu per kuartal. Dalam *query* pada gambar diatas menggunakan perintah *select* yang akan menampilkan data *borough*, data *sale_price* dengan penghitungan *average* yang diberi nama *average_sale_price*. Lalu menggunakan *case* sebagai pemilihan rentang data yang diubah kedalam kuartal yang diberi nama *sale_date*. Menggunakan *order by 2 desc* untuk mengurutkan data yang akan ditampilkan dari yang terbesar ke terkecil berdasarkan *average_sale_price*. Hasil dari menjalankan *query* tersebut yaitu menampilkan rata-rata penjualan dari tiap *borough* per kuartal.

Melakukan pencarian total penjualan berdasarkan data *building_class_category*. Dalam *query* pada gambar diatas menggunakan perintah *select* yang akan menampilkan data *building_class_category*, data *sale_price* dengan penghitungan *sum* yang diberi nama *total_sales_price*. Menggunakan *order by 2 desc* untuk mengurutkan data yang akan ditampilkan dari yang terbesar ke terkecil berdasarkan *total_sales_price*. Hasil dari menjalankan *query* tersebut yaitu menampilkan total penjualan dari kategori kelas bangunan yang diurutkan dari yang terbesar ke yang terkecil.

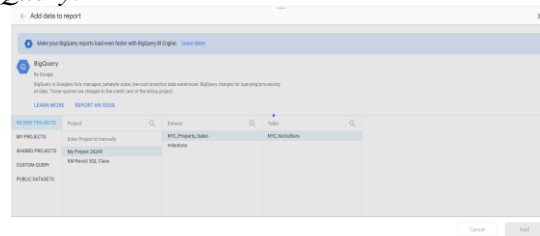
4. Visualisasi Data (*Visualization*)

4.1 Data Visualisasi (*Visualization*)

Menurut Sita (2022), Data visualisasi merupakan cara yang digunakan dalam mengubah suatu data atau informasi menjadi bentuk yang dapat lebih mudah dimengerti oleh orang lain dengan membuatnya ke dalam objek visual seperti gambar grafik, titik, garis, batang dan lainnya. Pada tahapan *visualization* dilakukan proses visualisasi data dari dataset yang telah dibersihkan sebelumnya dan sebagai lanjutan dari tahapan EDA menggunakan *Looker Studio*. Dalam visualisasi data terdapat beberapa tahapan yang dilalui dalam prosesnya, yaitu penghubungan data ke *Looker Studio*, pembuatan *scorecard*, pembuatan *chart* dan tabel, pembuatan *filter*, dan *dashboard*.

4.2 Penghubungan Data Ke *Looker Studio*

Untuk melakukan tahapan proses visualisasi diperlukan penghubungan data yang sebelumnya telah diimport ke *Google Sheet* lalu dihubungkan ke dalam *BigQuery*. Pada *Looker Studio* data dihubungkan dengan *BigQuery*.



Gambar 9. Penghubungan Data *BigQuery* Ke *Looker Studio*

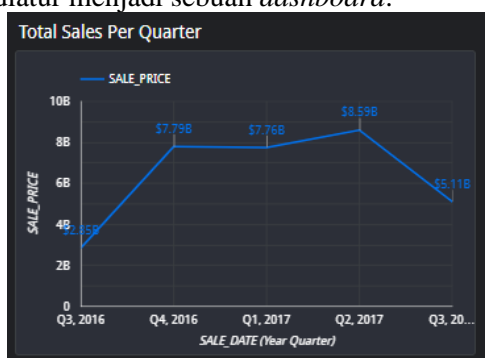
Untuk menghubungkan data yang ada pada *BigQuery* ke dalam *Looker Studio* dengan memilih *BigQuery* pada menu *Add Data*. Lalu memilih project, dataset serta tabel yang ingin dihubungkan ke dalam *Looker Studio*.

4.3 Pembuatan Scorecard

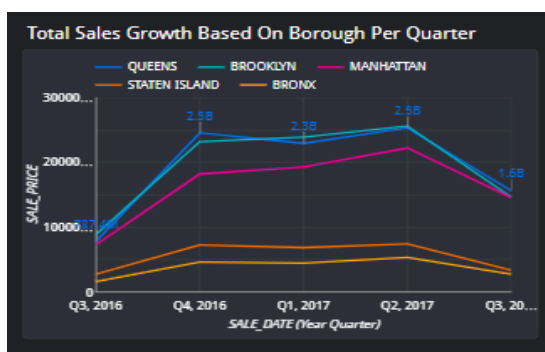
Dilakukan pembuatan *scorecard* yang berfungsi dalam menampilkan keseluruhan data yang dipilih untuk ditampilkan dalam *dashboard* dari dataset NYC Property Sales. Untuk menambahkan *Scorecard* pada *dashboard* yang ada pada menu *Add a chart* lalu memilih pilihan *scorecard compact number* yang ada pada pilihan sebelah kanan. Dalam *scorecard* ini akan digunakan untuk menampilkan *Total Sales*, *Average Sales*, dan *Total Unit Sales* secara keseluruhan data dengan *sale_price* sebagai data utamanya. Pada *scoreboard total sales* memilih data *sale_price* yang diubah namanya menjadi *Total Sales* dengan penghitungan *Sum* sebagai metric. Pada *Scoreboard Average Sales* memilih data *sale_price* yang diubah namanya menjadi *Average Sales* dengan penghitungan *average* sebagai metric. Pada *scoreboard total transaction* memilih data *sale_price* yang diubah namanya menjadi *Total Transaction* dengan penghitungan *count* sebagai metric.

4.4 Pembuatan Chart Dan Tabel

Pada langkah selanjutnya melakukan pembuatan *chart* dan tabel yang isinya merupakan visualisasi dari data yang telah di eksplorasi pada tahapan EDA. *Chart* dan tabel yang dibuat akan diatur menjadi sebuah *dashboard*.

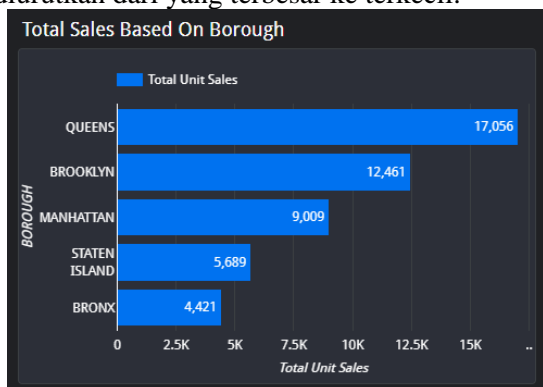


Gambar 10. Total Sales Per Quarter

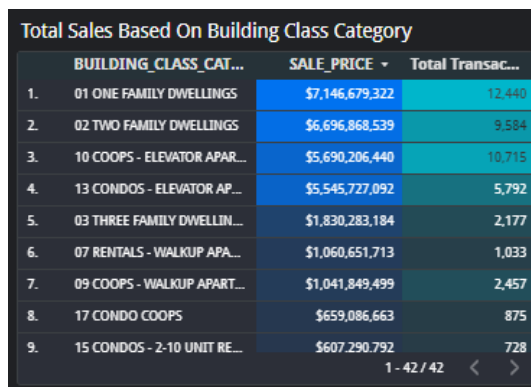


Gambar 11. Total Sales Based On Borough Per Quarter Chart

Untuk membuat chart pada gambar 10 dengan menggunakan menu *Add a Chart* lalu memilih *Line chart*. Mengatur isi dari *chart* tersebut dengan memilih data *Sale_Date* yang diatur tipenya ke dalam *year quarter* sebagai *dimension*, dan data *sale_price* dengan penghitungan menggunakan *sum* sebagai metric. Untuk membuat chart pada gambar 11 dengan menggunakan menu *Add a Chart* lalu memilih *Line chart*. Mengatur isi dari *chart* tersebut dengan memilih data *Sale_Date* yang diatur tipenya ke dalam *year quarter* sebagai *dimension*, data *Borough* sebagai *breakdown dimension*, dan data *sale_price* dengan penghitungan menggunakan *sum* sebagai metric. Selanjutnya memilih *sale_price* sebagai *dimension sort* dan memilih opsi *descending* agar data diurutkan dari yang terbesar ke terkecil.



Gambar 12. Total Transaction Based On Borough Chart



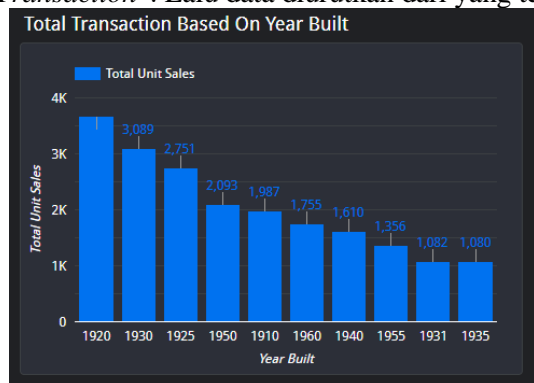
BUILDING_CLASS_CAT...	SALE_PRICE	Total Transac...
1. 01 ONE FAMILY DWELLINGS	\$7,146,679,322	12,440
2. 02 TWO FAMILY DWELLINGS	\$6,696,868,539	9,584
3. 10 COOPS - ELEVATOR APAR...	\$5,690,206,440	10,715
4. 13 CONDOS - ELEVATOR AP...	\$5,545,727,092	5,792
5. 03 THREE FAMILY DWELLIN...	\$1,830,283,184	2,177
6. 07 RENTALS - WALKUP APA...	\$1,060,651,713	1,033
7. 09 COOPS - WALKUP APART...	\$1,041,849,499	2,457
8. 17 CONDO COOPS	\$659,086,663	875
9. 15 CONDOS - 2-10 UNIT RE...	\$607,290,792	728

Gambar 13. Total Sales Based On Building Class Category Chart

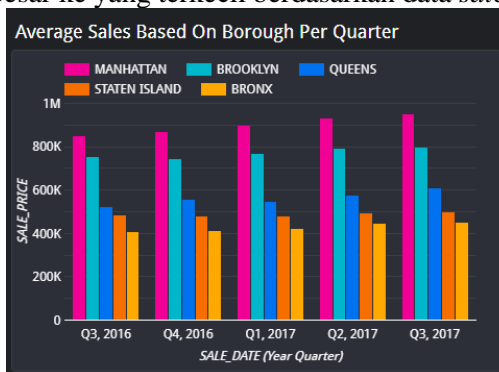
Membuat *chart* pada gambar 12 dengan menggunakan menu *Add a Chart* lalu memilih *Bar Chart*. Mengatur isi dari *chart* tersebut dengan memilih data *borough* sebagai *dimension*, data

sale_price sebagai metric dengan penghitungan *count* yang namanya diubah menjadi “*Total Unit Sales*”. Lalu di urutkan dari yang terbesar ke terkecil berdasarkan data *sale_price*.

Membuat chart pada gambar 13 dengan menggunakan menu *Add a Chart* lalu memilih *Table With Heatmap*. Mengatur isi dari tabel tersebut dengan memilih data *Building_class_category* sebagai *dimension*, memilih data *sale_price* dengan penghitungan *sum* sebagai *metric*, dan memilih data *sale_price* dengan penghitungan *count* sebagai *metric* lalu mengubah namanya menjadi “*Total Transaction*”. Lalu data diurutkan dari yang terbesar ke yang terkecil berdasarkan data *sale_price*.



Gambar 14. Total Transaction Based On Year Built



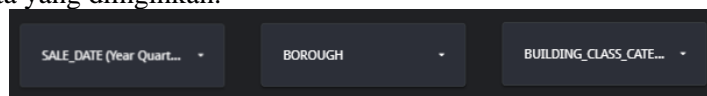
Gambar 15. Average Sales Based On Borough Per Quarter

Membuat chart pada gambar 14 dengan menggunakan menu *Add a Chart* lalu memilih *Column Chart* pada bagian Bar. Mengatur isi dari chart tersebut dengan memilih data *year_build* sebagai *dimension* dan memilih *sale_price* dengan penghitungan *count* yang diubah namanya menjadi “*Total Transaction*” sebagai *metric*. Lalu data diurutkan dari yang terbesar ke terkecil berdasarkan data total transaction.

Membuat chart pada gambar 15 dengan menggunakan menu *Add a Chart* lalu memilih *Column Chart* pada bagian Bar. Mengatur isi dari chart tersebut dengan memilih data *sale_date* sebagai *dimension* dengan penghitungan *Year Quarter*, memilih data *borough* pada bagian *breakdown dimension* dan memilih *sale_price* dengan penghitungan *average* sebagai *metric*. Lalu diurutkan berdasarkan data *sale_date* dengan urutan dari waktu yang lebih lama terlebih dahulu yaitu Q3 2016 dan mengurutkan rata-rata pada data *sale_price* dari data yang terbesar ke data yang terkecil.

4.5 Pembuatan Filter

Setelah melakukan pembuatan *scorecard*, chart dan tabel. Dilakukan pembuatan *filter* yang akan menampilkan data yang ada berdasarkan filter yang dipilih. Sehingga memudahkan pengguna dalam mencari data yang diinginkan.

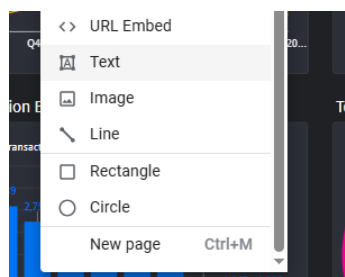


Gambar 16. Penambahan Filter

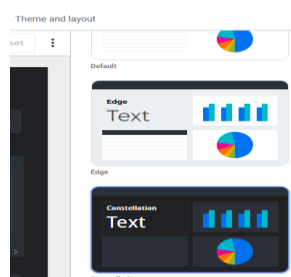
Untuk menambahkan filter berdasarkan data *sale_date* dengan penghitungan *year quarter*, *borough*, dan *building_class_category*. Filter dapat ditambahkan pada bagian *Add a control* lalu memilih *Drop-down list*. Pada filter *sale_date* memilih data *sale_date* dengan penghitungan *year quarter* pada bagian *control field*. Pada *filter borough* memilih data *borough* pada bagian *control field*. Pada *filter building_class_category* memilih data *building_class_category* pada bagian *control field*. Semua filter menggunakan data *sale_price* sebagai *metric*.

4.6 Dashboard

Setelah melalui semua tahapan yang ada, maka data yang ingin di visualisasikan telah berhasil ditampilkan. Selanjutnya adalah merapikan tata letak dari dashboard yang ingin dibuat. Untuk menunjukkan kepada para pengguna mengenai *dashboard* yang ditampilkan maka ditambahkan judul di bagian atas dashboard.



Gambar 17. Menambahkan Judul



Gambar 18. Mengubah Tema

Untuk menambahkan judul pada *dashboard* (gambar 17) dengan memilih insert lalu text. Setelah itu letakkan kursor di bagian atas *dashboard* untuk meletakkan judul dashboard. Klik dua kali pada bagian teks yang sudah ditambahkan untuk menuliskan judul dashboard. Judul dashboard ditulis “NYC PROPERTY SALES (2016-2017). Dengan penulisan judul ini pengguna dapat dengan jelas melihat bahwa dashboard yang ada menampilkan data dari NYC *Property Sales* dengan rentang waktu 2016 sampai 2017.

Melakukan perubahan tema (gambar 18) dengan memilih tema yang tersedia pada *Looker Studio*. Perubahan dilakukan dengan mengklik menu *Theme and Layout*, lalu memilih tema yang gelap seperti pada gambar diatas. Perubahan ini telah dilakukan pada tahapan awal visualisasi data.



Gambar 19. Dashboard NYC Property Sales

Gambar 19 merupakan tampilan akhir dari *dashboard NYC Property Sales* yang telah melalui tahap-tahap yang telah dikerjakan sebelumnya. Dengan dashboard ini pengguna diharapkan dapat mengakses informasi yang ada dengan nyaman dan mudah dimengerti. Dashboard ini dapat diakses dengan menggunakan platform *Looker Studio* melalui tautan berikut:
<https://lookerstudio.google.com/reporting/a60afdae-b109-4043-99dc-1fa3f4b3df9f>.

PENUTUP

Adapun kesimpulan yang didapat dari penulisan ilmiah yang dilakukan, berhasil membuat visualisasi data yang merupakan *dashboard* yang interaktif sebagai hasil akhirnya dengan menggunakan *Looker Studio* dalam menganalisis dataset NYC *Property Sales*. Dashboard interaktif yang dibuat memberikan informasi terkait tren waktu pada penjualan properti, rata-rata harga properti, penyebaran distribusi penjualan dan transaksi penjualan dalam suatu *borough*, kelas bangunan ataupun tahun pembuatan bangunan properti. Informasi yang didapat dari dashboard yang berhasil dibuat yaitu waktu dengan peningkatan penjualan terdapat pada kuartal 3, *borough Queens* dengan *Brooklyn* sebagai borough yang dominan, mempertimbangkan kategori bangunan *One Family Dwellings* dan usia bangunan yang lebih tua, khususnya tahun 1920. Hal ini dapat memberikan informasi bagi para investor untuk melakukan investasi di waktu yang tepat serta memberikan wawasan yang berharga. *Dashboard* dapat dilihat dan digunakan dengan mengakses platform *Looker Studio* melalui tautan berikut ini:

<https://lookerstudio.google.com/reporting/a60afdae-b109-4043-99dc-1fa3f4b3df9f/page/bQ64D> .
Dalam pengembangan dashboard visualisasi data yang telah dibuat dapat ditambahkan fitur-fitur dan chart baru seperti *map chart* yang berguna dalam memvisualisasikan distribusi di seluruh

wilayah geografis, sehingga memberikan pemahaman yang jelas dan intuitif tentang tren dan pola regional. Dan juga dapat menerapkan *dashboard* dalam suatu *website*.

REFERENSI

- Anak Agung Aryasatya Daniswara, I Kadek Dwi Nuryana. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. Vol 5, No 1 2023. <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/55395/43970>.
- Darmanto, Y., Koryanto, L., & Faizah, N. (2023). Perancangan Aplikasi Visualisasi Database Produksi dengan Metode Agregasi menggunakan Looker Studio dan Google Sheets Berbasis Web Studi Kasus di PT. ANTAM. Tbk. *Journal Innovations Computer Science*, 2(2), 50–62. <https://doi.org/10.56347/jics.v2i2.141>
- Fauzan Alghifari, Didi Juardi. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. <https://ejournal.upbatam.ac.id/index.php/jif/article/view/3755/2062>
- Feliks Victor Parningotan Samosir, Loudry Palmarums Mustamu, Erik Dwi Anggara, Albertus Indarko Wiyogo, Andreas Widjaja. (2021). Exploratory Data Analysis Terhadap Kepadatan Penumpang Kereta Rel Listrik. Vol 7, No 2 2021. http://dx.doi.org/10.28932/jutisi.v7i2.3700_
- Fitri Ariani, A., Aulia, K., & Ahmad Arafat, L. O. (2024). Pengembangan Dashboard Interaktif Menggunakan Looker Studio Untuk Visualisasi Dan Prediksi Harga Komoditas Cabe Di Jawa Timur. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 8067–8074. <https://doi.org/10.36040/jati.v8i4.10616>
- Irmayani, W. (2021). Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi. *Jurnal Khatulistiwa Informatika*, IX(1), 68–72.
- Junaidi. (2014). Deskripsi Data Melalui Box Plot. Fakultas Ekonomi dan Bisnis Universitas Jambi. <http://dx.doi.org/10.13140/RG.2.1.1057.8088> .
- Nisak Ruwah Ibnatur Husnul. (2020). Statistik Deskriptif. https://repository.unpam.ac.id/8670/1/SAK0233_Statistik_Deskriptif-full.pdf .
- Sita Muharni, Apri Candra. (2022). Buku Visualisasi Data Menggunakan Data Studio. https://www.researchgate.net/publication/362791707_Buku_Visualisasi_Data_Menggunakan_Data_Studio_
- Yahya, & Mahpuz. (2019). Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat. *Infotek : Jurnal Informatika Dan Teknologi*, 2(2), 1–23.
- Yeni Setiani, Nabila Rachmah, Indra Purnama. (2023). Visualisasi Data Malnutrisi Anak Di Asia Menggunakan Looker Studio Serta Analisis Data Dengan Metode ANOVA. Vol 3, No 3 2023, Hal 188-212. <https://doi.org/10.55606/juisik.v3i3.701> .