

EKSPLORATIF DATA ANALISIS TERHADAP DATA WORLD POPULATION

Rizki Amanda¹, Adelia Yuandhika², Nezalfa Sabrina³,
Jasmine Aulia⁴, Cahya Eka Melati⁵

^{1,2,3,4,5} Jurusan Sains Data, UPN “Veteran” Jawa Timur

¹22083010045@student.upnjatim.ac.id, ²22083010066@student.upnjatim.ac.id, ³22083010067@student.upnjatim.ac.id

⁴22083010074@student.upnjatim.ac.id, ⁵22083010090@student.upnjatim.ac.id

Abstract: World population growth is a crucial issue that requires thorough examination. The World Population Prospects 2022 report by the United Nations (UN) projects that the global population will reach 10.4 billion people by the 2080s. This projection indicates that world population growth will continue but at a progressively slower rate. This study aims to conduct exploratory data analysis on the projected world population growth. Exploratory data analysis was performed using simple arithmetic techniques and graphical methods to understand the characteristics of population growth and influencing factors. Based on the data analysis, several conclusions can be drawn. Firstly, the continents of Africa, Asia, and Europe exhibit significant population growth compared to others. Secondly, Asia had the highest population in 2022, followed by Africa and Europe. Thirdly, there is no strong correlation between land area and population growth. Fourthly, some countries with small land areas can have high population densities. Lastly, a linear regression model can predict population growth, but the accuracy of the model needs enhancement. Overall, this analysis provides deep insights into population growth characteristics and influencing factors. It can be utilized to support decision-making and the development of population growth-related strategies.

Keywords: population growth, exploratory data analysis, linear regression model

Abstrak: Pertumbuhan populasi dunia merupakan salah satu isu penting yang perlu dikaji. Laporan Prospek Populasi Dunia 2022 dari Perserikatan Bangsa-Bangsa (PBB) memproyeksikan bahwa populasi dunia akan mencapai 10,4 miliar jiwa pada tahun 2080-an. Proyeksi ini menunjukkan bahwa pertumbuhan populasi dunia akan terus meningkat, tetapi dengan laju yang semakin melambat. Penelitian ini bertujuan untuk melakukan analisis data eksploratif terhadap proyeksi pertumbuhan populasi dunia. Analisis data eksploratif dilakukan menggunakan teknik aritmatika sederhana dan teknik grafis untuk memahami karakteristik pertumbuhan populasi dan faktor-faktor yang dapat memengaruhinya. Berdasarkan analisis data yang dilakukan, dapat ditarik beberapa kesimpulan. Pertama, benua Afrika, Asia, dan Eropa memiliki tingkat pertumbuhan populasi yang signifikan dibandingkan dengan benua lainnya. Kedua, benua Asia memiliki populasi tertinggi pada tahun 2022, diikuti oleh Afrika dan Eropa. Ketiga, tidak ada korelasi yang kuat antara luas wilayah dan pertumbuhan populasi. Keempat, beberapa negara dengan luas wilayah kecil dapat memiliki kepadatan penduduk yang tinggi. Kelima, model regresi linier dapat digunakan untuk memprediksi pertumbuhan populasi, tetapi akurasi model tersebut perlu ditingkatkan. Secara keseluruhan, analisis ini memberikan wawasan mendalam tentang karakteristik pertumbuhan populasi dan faktor-faktor yang dapat memengaruhinya. Analisis ini dapat digunakan untuk mendukung pengambilan keputusan dan pengembangan strategi terkait pertumbuhan populasi.

Kata kunci: pertumbuhan populasi, analisis data eksploratif, model regresi linier

I. PENDAHULUAN

1. LATAR BELAKANG

Penelitian ini bertujuan untuk menyelidiki proyeksi pertumbuhan populasi dunia yang telah dirilis oleh Perserikatan Bangsa-Bangsa (PBB) pada tahun 2022. Pada tanggal 15 November 2022, laporan Prospek Populasi Dunia 2022 yang dirilis oleh PBB memproyeksikan

peningkatan populasi dunia menjadi 8 miliar jiwa. Salah satu proyeksi yang mencolok adalah prediksi bahwa India akan melampaui Cina sebagai negara terpadat pada tahun 2023. Proyeksi lanjutan PBB mengindikasikan tren pertumbuhan yang terus meningkat, dengan prediksi mencapai

8,5 miliar jiwa pada tahun 2030 dan 9,7 miliar pada tahun 2050. Titik puncak pertumbuhan ini diperkirakan terjadi pada tahun 2080-an, memuncak di sekitar 10,4 miliar jiwa, dan diprediksi akan stabil hingga tahun 2100. Meskipun pertumbuhan populasi dunia mengalami perlambatan pada tingkat terendah sejak tahun 1950, dengan penurunan menjadi kurang dari satu persen pada tahun 2020, laporan ini menggarisbawahi penurunan yang signifikan dalam tingkat kesuburan dan dampaknya terhadap pertumbuhan populasi. Data menunjukkan bahwa dua pertiga dari populasi global saat ini tinggal di negara atau wilayah dengan tingkat fertilitas di bawah 2,1 kelahiran per wanita, yang merupakan tingkat yang dianggap perlu untuk mencapai pertumbuhan nol dalam jangka panjang.

Laporan PBB juga mengidentifikasi 61 negara yang diprediksi akan mengalami penurunan populasi setidaknya satu persen dalam tiga dekade mendatang. Penurunan ini dipicu oleh tingkat kesuburan yang rendah dan peningkatan tingkat emigrasi. Pandemi COVID-19 juga menjadi sorotan dalam laporan ini, dengan catatan penurunan harapan hidup global menjadi 71 tahun pada tahun 2021 dari 72,9 tahun pada tahun 2019. Pandemi ini diyakini telah menyebabkan penurunan jangka pendek dalam jumlah kehamilan dan kelahiran di beberapa negara. John Wilmoth, Direktur Divisi Kependudukan PBB Departemen

Ekonomi dan Sosial (DESA), menyoroti bahwa tindakan pemerintah untuk mengurangi kesuburan mungkin memiliki dampak kecil pada laju pertumbuhan penduduk saat ini. Namun, akumulasi dari kesuburan yang lebih rendah dapat menghasilkan perlambatan yang substansial pada pertumbuhan populasi global pada paruh kedua abad ini, sesuai dengan analisis dari United Nations | Population Division [1].

Laporan ini akan mengupas lebih dalam isu-isu yang terkait dengan proyeksi pertumbuhan populasi dunia, menyelidiki penelitian sebelumnya yang relevan dengan konteks saat ini, serta menggali aspek-aspek yang perlu diperhatikan dalam memahami dinamika pertumbuhan populasi global.

2. TINJAUAN PUSTAKA

2.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah suatu metode eksplorasi data yang menggunakan teknik aritmatika sederhana dan teknik grafis untuk merangkum data pengamatan. Eksplorasi data memiliki peran integral dalam pemahaman kita terhadap data, memberikan dukungan yang signifikan dalam mengeksplorasi dan menemukan karakteristik-karakteristik data yang nantinya dapat digunakan untuk memilih model statistik yang sesuai [2]. Meskipun EDA banyak menggunakan teknik grafis, tetapi tidak identik dengan

analisis grafik. Pendekatan analisis secara grafis lebih terbatas pada perangkat teknik grafis, sedangkan EDA mencakup area yang lebih luas dan menekankan pendekatan langsung sehingga data sendiri yang mengungkap struktur dan modelnya [3].

2.2 Anaconda

Anaconda adalah sebuah aplikasi open source gratis yang berfungsi untuk distribusi Python dan R. Python banyak digunakan dalam berbagai perhitungan ilmiah, termasuk dalam bidang machine learning, pengolahan data dalam skala besar, analisis prediksi, dan banyak lagi. Aplikasi ini menyimpan lebih dari 1500 package populer yang dapat diakses di beragam platform sistem operasi seperti Windows, Linux, dan MacOS [4].

Anaconda memiliki banyak kelebihan diantaranya yaitu mudah dan cepat dalam menginstall package untuk data science, pengguna dapat memilih versi bahasa pemrograman python yang ingin diaplikasikan, lalu library dependensi dan lingkungan juga dapat diatur menggunakan Anaconda. Namun Anaconda juga memiliki kekurangan yaitu pengguna wajib mempunyai pengalaman luas dengan bahasa

pemrograman Python, lalu besarnya file installer yang mencapai ratusan MB dapat memakan banyak ruang laptop atau PC pengguna. [5].



Gambar 1. Anaconda Logo

2.3 Python 3

Python adalah bahasa pemrograman yang umumnya digunakan dalam pengembangan aplikasi berbasis web. Secara umum, logika perintah yang ditulis dalam Python lebih mudah digunakan dibandingkan dengan bahasa pemrograman tingkat rendah seperti bahasa assembly [6]. Keberadaan perpustakaan dan pustaka standar membuat Python menjadi pilihan yang kuat untuk pengembangan perangkat lunak [7].



Gambar 2. Python 3 Logo

3. PENELITIAN TERKAIT

Dalam penelitian "Stemplot" tahun 2023, Ade Setiawan menerapkan analisis data eksploratif menggunakan Stem-and-

Leaf Plot. Metode ini memisahkan setiap nilai menjadi bagian batang dan daun, memberikan gambaran visual jelas tentang distribusi data. Tujuannya adalah memperkenalkan dan menjelaskan

penggunaan Stem-and-Leaf Plot dalam analisis data eksploratif, memberikan panduan bagi peneliti untuk memahami dan menggambarkan dataset secara efektif melalui visualisasi tersebut.

Dalam penelitian "Deskripsi Data Melalui Box-Plot" (Junaidi, 2015), analisis data eksploratif menggunakan boxplot untuk memahami karakteristik dan distribusi data numeris. Boxplot memberikan representasi visual komprehensif terhadap statistik deskriptif kunci tanpa asumsi distribusi tertentu. Metode ini fleksibel dan dapat menunjukkan perbedaan antara populasi, serta mencerminkan derajat dispersi dan skewness data. Penelitian ini memberikan kontribusi pada pemahaman dan

penggunaan efektif boxplot dalam menganalisis dan mendeskripsikan data numeris secara visual.

Dalam penelitian "Deforestation and World Population Sustainability" (Bologna & Aquino, 2020) menggunakan analisis kuantitatif untuk mengevaluasi keberlanjutan pertumbuhan populasi dunia terkait deforestasi. Studi ini menyimpulkan bahwa kemungkinan keruntuhan populasi manusia akibat konsumsi sumber daya sangat mungkin, dengan probabilitas bertahan kurang dari 10%. Perubahan besar dalam parameter seperti pertumbuhan populasi dan laju deforestasi diperlukan untuk menghindari skenario tersebut, menyoroti urgensi tindakan kolektif untuk menjaga keberlanjutan peradaban.

II. METODE PENELITIAN



Gambar 3. *Desain Sistem Analisis Data Eksploratif*

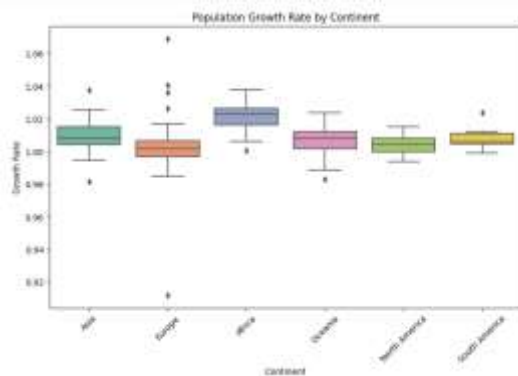
Desain sistem kami dirancang dengan urutan langkah-langkah yang terstruktur untuk memahami dan menggali wawasan dari dataset. Tahap pertama

adalah Data Presentation, di mana data ditampilkan dalam bentuk grafik, tabel, atau visualisasi lainnya untuk memberikan pemahaman awal. Langkah berikutnya adalah Data Transformation and Analysis, yang melibatkan pembersihan data, transformasi, dan analisis awal untuk mengidentifikasi pola atau kelompok

dalam data. Setelah itu, Descriptive Statistic digunakan untuk menghitung rata-rata, median, modus, dan standar deviasi, serta membuat visualisasi seperti histogram atau boxplot pada dataset. Terakhir, Regression Testing diterapkan untuk mengidentifikasi hubungan antara variabel-variabel dalam dataset melalui penerapan model regresi, uji signifikansi, dan evaluasi

III. HASIL DAN PEMBAHASAN

3.1 Data presentation

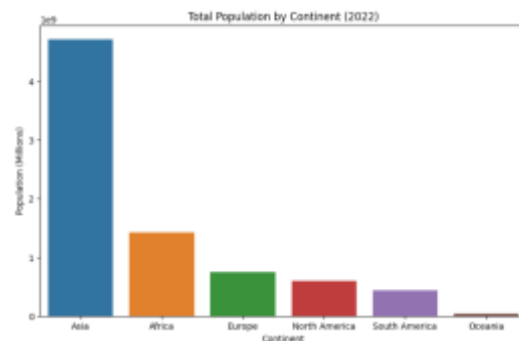


Gambar 4. Population Growth Rate by Continent

Berdasarkan diagram boxplot diatas, nilai tengah pada benua Afrika, Asia, dan Eropa memiliki nilai tengah lebih tinggi dan memiliki penyebaran data pada bagian atas yang lebih luas dari pada benua-benua lain. Oleh karena itu, negara - negara pada benua Afrika, Asia, dan Eropa memiliki tingkat pertumbuhan yang jauh lebih tinggi daripada negara-negara di benua lainnya. Sebaliknya pada benua Oseania dan Amerika Selatan memiliki nilai tengah lebih rendah dan distribusi data yang lebih condong ke bagian bawah, yang

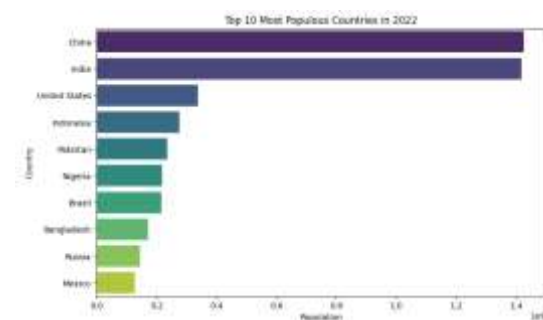
kinerja model. Dengan langkah-langkah ini, analis data dapat dilakukan dari eksplorasi awal hingga analisis mendalam, memberikan wawasan yang diperlukan untuk pengambilan keputusan dan pengembangan strategi analisis lebih lanjut. Pemanfaatan visualisasi data juga mendukung interpretasi hasil analisis dengan lebih baik.

artinya menunjukkan bahwa pertumbuhan populasi pada benua tersebut lebih lambat.



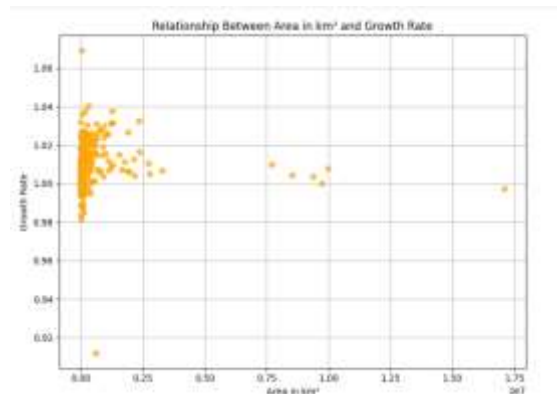
Gambar 5. Total Population by Continent (2022)

Berdasarkan diagram batang tersebut, benua Asia memiliki total populasi terbanyak pada tahun 2022, disusul oleh benua Afrika dan Eropa. Sebaliknya benua Oseania memiliki total populasi paling sedikit dari benua lainnya.



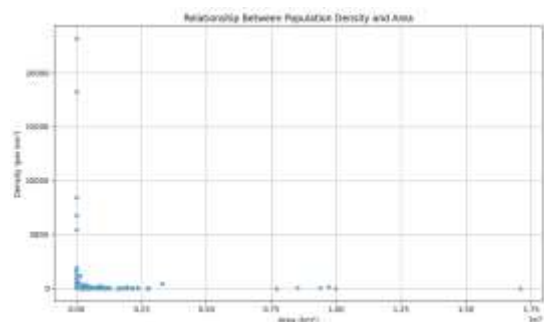
Gambar 6. Top 10 Most Populous Countries in 2022

Berdasarkan diagram batang tersebut, dapat dilihat bahwa negara China memiliki populasi terbesar pada tahun 2022, disusul oleh negara India. Dapat dilihat pada output diatas negara Indonesia menduduki peringkat keempat pada tahun 2022, satu tingkat dibawah Amerika Selatan.



Gambar 7. Relations Between Area in km^2 and Growth Rate

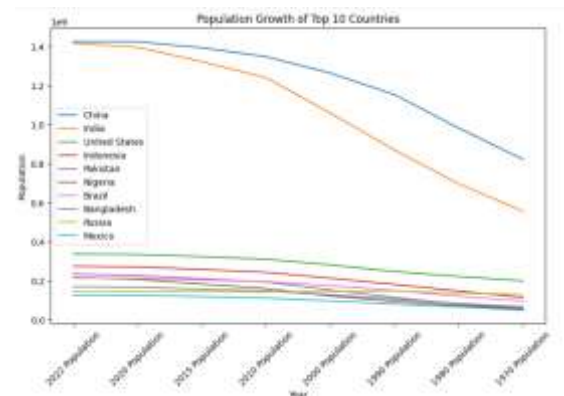
Berdasarkan scatter plot tersebut tidak memiliki pola tertentu, yang mana menunjukkan bahwa tidak ada korelasi yang kuat antara luas wilayah dan tingkat pertumbuhan populasi.



Gambar 8. Relations Between Population Density and Area

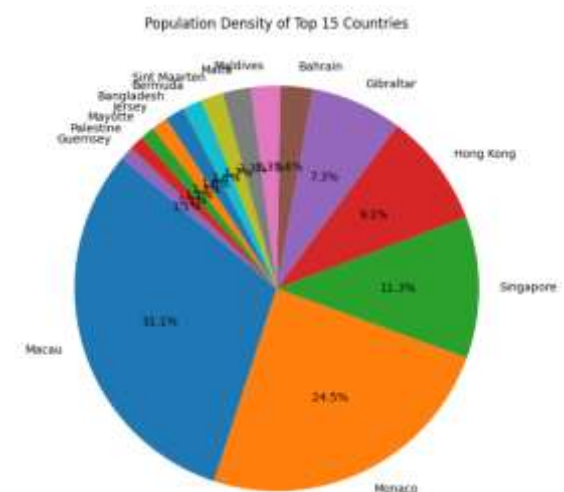
Berdasarkan hasil output tersebut, titik-titik dalam scatter plot berkelompok. Ini menunjukkan adanya hubungan antara kedua variabel. Artinya terdapat

kemungkinan beberapa negara-negara dengan luas wilayah yang kecil memiliki kepadatan penduduk yang tinggi.



Gambar 9. Population Growth of Top 10 Countries

Berdasarkan diagram garis diatas, dapat dilihat 10 negara dengan populasi terbesar sama-sama menunjukkan kenaikan populasi dari tahun 1990 sampai 2022. Dapat diambil kesimpulan bahwa pertumbuhan populasi konsisten naik dari waktu-waktu.



Gambar 10. Population Density of Top 15 Countries

Berdasarkan diagram lingkaran tersebut, dapat dilihat gambaran proporsi kepadatan 15 negara atau wilayah dengan kepadatan tertinggi. Persentase kepadatan

tertinggi diperoleh Macau (China) dengan persentase sebesar 31,1%, disusul oleh negara Monaco dengan persentase sebesar 24,5%, dan Singapore dengan persentase sebesar 11,3%.

3.2 Data Transformation and Analysis

3.2.1 Cek Duplikasi Data

```
duplikasi = df[df.duplicated()]
if not duplikasi.empty:
    print("Duplikasi Data:")
    print(duplikasi)
else:
    print("Tidak Ada Duplikasi Dalam Dataset")
```

```
Tidak Ada Duplikasi Dalam Dataset
```

Kode pemrograman python seperti yang tertera di atas untuk mengetahui apakah terdapat duplikasi data di dalam dataset yang digunakan. Hasilnya dapat dilihat bahwa dataset tidak memiliki data yang terduplikasi.

3.2.2 Cek Missing Value

```
missing_values = df.isnull().sum()
print("Kolom Dengan Missing Values :")
print(missing_values[missing_values > 0])
```

```
Kolom Dengan Missing Values :
Series([], dtype: int64)
```

Pengecekan terhadap ada tidaknya missing values dalam dataset menggunakan kode skrip di atas. Setelah menjalankan kode skrip tersebut dapat diketahui bahwa dataset tidak memiliki missing values. Jadi dataset tersebut dapat langsung diproses untuk langkah selanjutnya.

3.2.3 Analisis Umum

```
Africa      57
Asia        50
Europe       50
North America 40
Oceania      23
South America 14
Name: Continent, dtype: int64
```

Disini dapat diketahui jumlah negara yang terdapat dalam masing-masing benua, yaitu Benua Afrika dengan 57 negara, Benua Asia dengan 50 negara, Benua Eropa dengan 50 negara, Amerika Utara dengan 40 negara, Oceania dengan 23 negara, dan Amerika Selatan dengan 14 negara.

```
Benua dengan negara terbanyak :
Africa
```

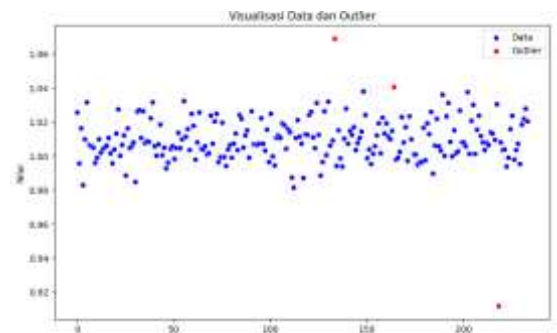
Menggunakan `idxmax()` untuk mengetahui benua yang memiliki negara terbanyak. Hasilnya adalah Benua Afrika sebagai benua yang memiliki negara terbanyak di dalamnya.

```
Benua dengan negara paling sedikit :
South America
```

Menggunakan `idxmin()` untuk mengetahui benua yang memiliki negara yang paling sedikit. Hasilnya adalah Amerika Selatan sebagai benua yang

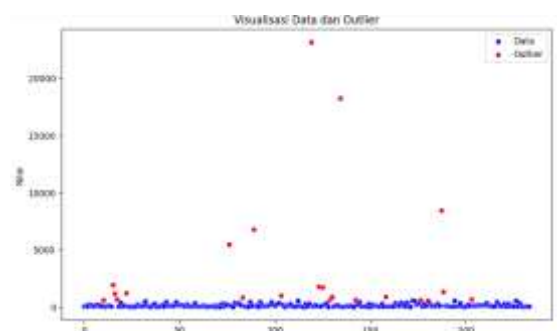
memiliki negara yang jumlahnya paling sedikit diantara benua lainnya.

3.2.4 Visualisasi dan Outlier



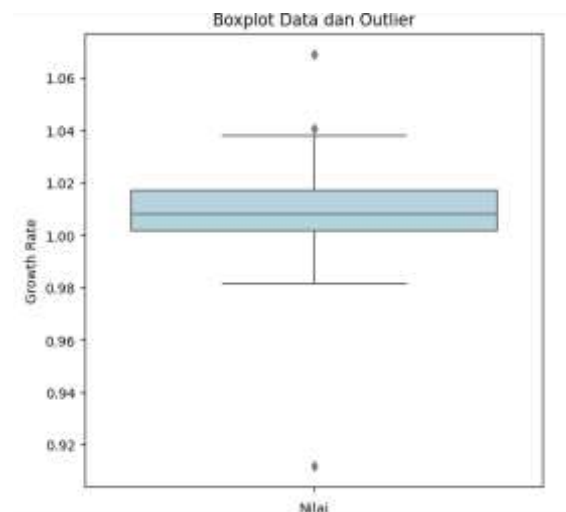
Gambar 11. Visualisasi Data dan Outlier

Berdasarkan output tersebut dapat diketahui bahwa data pada kolom Growth Rate memiliki 3 outlier yang ditandai dengan scatterplot yang berwarna merah. Sementara untuk kolom density per km hasilnya seperti gambar berikut ini :



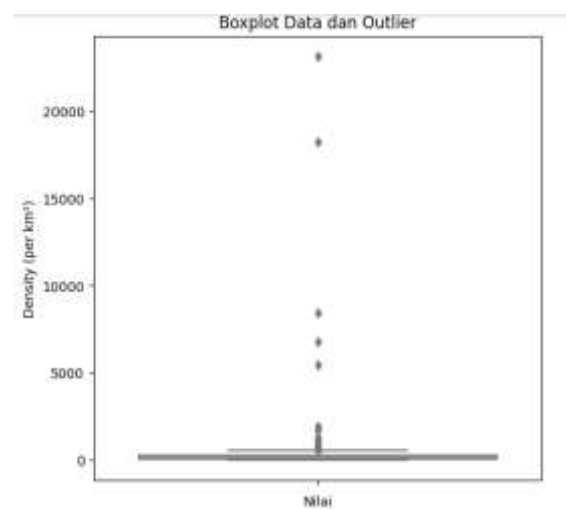
Gambar 12. Visualisasi Data dan Outlier

Apabila menggunakan boxplot untuk kolom Growth Rate maka visualisasinya akan menjadi seperti di bawah ini.

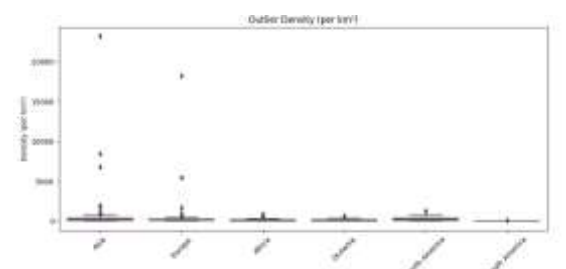


Gambar 13. Boxplot Data dan Outlier

Sedangkan untuk kolom Density per KM adalah sebagai berikut :



Gambar 14. Boxplot Data dan Outlier



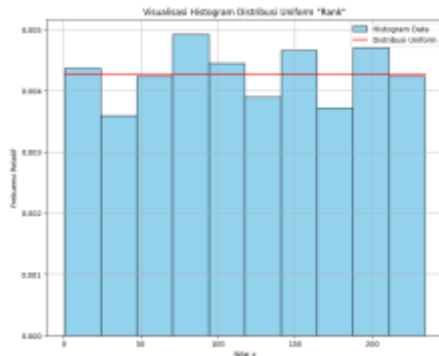
Gambar 15. Outlier Density (per km²)

Dapat diketahui bahwa data memiliki sebagian kecil outlier. Oleh sebab

itu hubungan antar variabel tidak akan terlalu terpengaruh, dan hasil uji statistik memiliki kemungkinan untuk tetap valid.

3.3 Descriptive Statistic

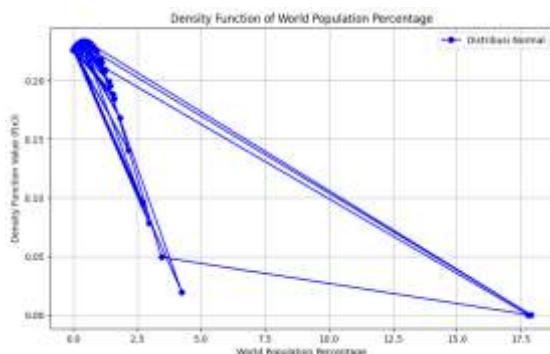
3.3.1 Uniform Distribution



Gambar 16. Visualisasi Histogram
Distribusi Uniform 'Rank'

Berdasarkan hasil output dari Distribusi Uniform menghasilkan nilai a sebesar 0.50 dan nilai b sebesar 234.50. Serta didapatkan nilai probabilitas $P(X < 5)$ sebesar 0.02 dengan persentase 1.92%. Dengan nilai sebesar 0.02 atau 2%, kita dapat mengatakan bahwa peluang mendapatkan nilai X yang kurang dari 5 relatif rendah. Persentase probabilitas 1.92% memberikan gambaran lebih intuitif tentang seberapa kecil peluang ini dalam skala persentase.

3.3.2 Normal Distribution



Gambar 17. Density Function of World Population Percentage

Berdasarkan hasil output Distribusi Normal, didapatkan dalam distribusi probabilitas kontinu, $f(x)$ menunjukkan seberapa tinggi kurva distribusi pada nilai tertentu. Semakin tinggi nilai $f(x)$, semakin tinggi kepadatan probabilitas di sekitar nilai x tersebut.

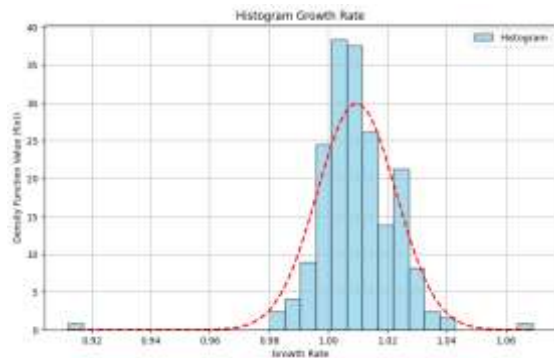
3.3.3 Skewness

x	f(x)
0	1.0257 14.414890
1	0.9957 17.410202
2	1.0164 26.215679
3	0.9831 4.186696
4	1.0100 29.854151
...	...
229	0.9953 16.869232
230	1.0184 24.014455
231	1.0217 19.785056
232	1.0280 11.537083
233	1.0204 21.510297

[234 rows x 2 columns]

Skewness: -1.09

Distribusi miring ke kiri (negatif skewness)



Gambar 18. Histogram Growth Rate

Berdasarkan output Skewness, didapatkan bahwa setiap baris dalam tabel mencantumkan nilai x (Growth Rate) dan nilai $f(x)$ yang merupakan nilai fungsi densitas probabilitas untuk nilai x tersebut dalam distribusi normal.

Histogram ditampilkan dalam bentuk batang-batang yang merepresentasikan frekuensi relatif dari setiap bin (interval) pada data pertumbuhan populasi. Jika histogram dan kurva distribusi normal cukup cocok, itu menunjukkan bahwa data pertumbuhan populasi mungkin memiliki distribusi yang mendekati distribusi normal. Jika kurva distribusi normal sesuai dengan histogram, distribusi pertumbuhan populasi mungkin tidak memiliki skewness yang signifikan dan mendekati bentuk simetris.

Nilai Skewness didapatkan dari $df["Growth Rate"]$ sebesar -1.09 yang berarti data berdistribusi miring ke kiri atau negative skewness. Dengan nilai skewness negatif, ekor distribusi (nilai-nilai di sebelah kiri) lebih panjang daripada di sebelah kanan. Mayoritas nilai cenderung berada di sebelah kanan nilai rata-rata.

3.3.4 Kurtosis

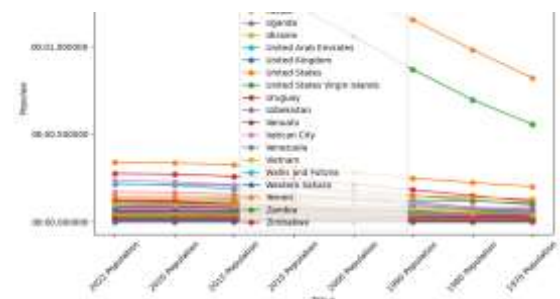
Kurtosis: 12.20

Distribusi leptokurtic (puncak lebih tajam, ekor lebih berat)

Berdasarkan hasil output Kurtosis, didapatkan nilai dari $df["Growth Rate"]$ sebesar 12.20 yang berarti data berdistribusi leptokurtic. Distribusi data cenderung memiliki nilai-nilai yang sangat terkonsentrasi di sekitar nilai rata-rata, sehingga puncaknya lebih tinggi dan lebih tajam. Ekor distribusi lebih berat, yang berarti adanya banyak nilai yang jauh dari rata-rata.

Distribusi ini memiliki lebih banyak "puncak" di sekitar nilai rata-rata, dan ekor distribusi lebih berat dibandingkan dengan distribusi normal. Fenomena ini disebut sebagai distribusi leptokurtic, yang menunjukkan adanya potensi untuk menghasilkan nilai ekstrim (outliers) atau kejadian yang jarang.

3.4 Regression Testing



Gambar 19. Visualisasi dalam beberapa tahun

Berdasarkan grafik diatas dapat diketahui kepadatan penduduk per kilometer, ini memberikan visual tentang bagaimana kepadatan penduduk berubah dari beberapa entitas.



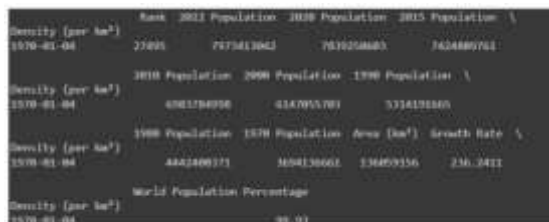
Gambar 20. Pengelompokan berdasarkan bulan

Berdasarkan output diatas dapat kita ketahui berdasarkan pengelompokan kepadatan penduduk per bulannya.



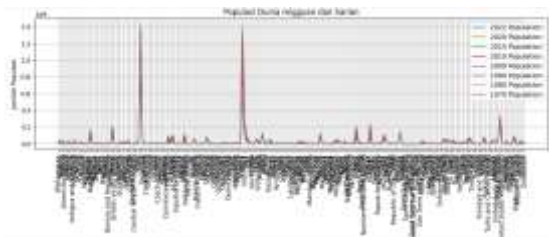
Gambar 21. Pengelompokan berdasarkan hari

Berdasarkan output diatas dapat kita ketahui berdasarkan pengelompokan kepadatan penduduk per harinya.



Gambar 22. Pengelompokan berdasarkan minggu

Berdasarkan output tersebut dapat kita ketahui berdasarkan pengelompokan kepadatan penduduk per minggunya.



Gambar 23. Visualisasi rata-rata mingguan dan harian

Berdasarkan plot diatas menampilkan visualisasi perubahan populasi hasil rata-rata per minggu dan hari.

Rata-rata dari kepadatan penduduk:

452.1270435897435

T-Statistic: 3.347437458358648

P-Value: 0.000475435603049331

Hipotesis Nol Ditolak. Ada cukup bukti bahwa rata-rata kepadatan penduduk lebih besar dari nilai yang diuji.

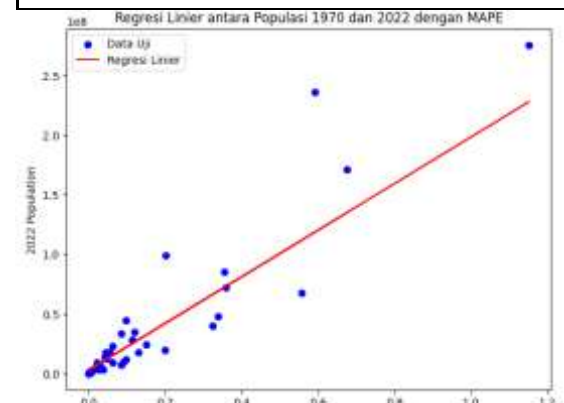
Dari hasil kode diatas kita mendapatkan nilai P-Value: 0.000475435603049331 karna P-Value cukup kecil dari tingkat signifikansi maka H_0 ditolak, ini menunjukkan ada cukup bukti bahwa rata-rata kepadatan penduduk lebih besar dari nilai yang diuji.

Nilai Korelasi Pearson: -

0.06312785050059672

Berdasarkan nilai korelasi pearson yang diperoleh sebesar -0.063 maka dapat diketahui bahwa dari 2 variabel, Density (per km²) dan Area (km²), tidak memiliki hubungan yang linier.

MAPE: 1860.7770643559452



Gambar 24. Regresi dan evaluasi

Plot diatas memvisualisasikan hubungan antara populasi pada tahun 1970 dan populasi pada tahun 2022, serta mendapatkan hasil MAPE sebesar 1860.7770643559452 sebagai ukuran evaluasi kinerja model regresi linier.

IV. KESIMPULAN

Berdasarkan analisis data yang dilakukan, dapat ditarik beberapa kesimpulan. Melalui diagram boxplot, terlihat bahwa benua Afrika, Asia, dan Eropa memiliki nilai tengah pertumbuhan populasi yang lebih tinggi dan penyebaran data yang lebih luas, menunjukkan tingkat pertumbuhan yang signifikan dibandingkan dengan benua lainnya. Benua Asia menunjukkan total populasi tertinggi pada tahun 2022, diikuti oleh Afrika dan Eropa, sementara Oseania memiliki populasi paling sedikit. Scatter plot menunjukkan

bahwa tidak ada korelasi yang kuat antara luas wilayah dan pertumbuhan populasi, tetapi beberapa negara dengan luas wilayah kecil dapat memiliki kepadatan penduduk yang tinggi. Diagram lingkaran memberikan gambaran proporsi kepadatan tertinggi, dengan Macau (China), Monaco, dan Singapura memiliki persentase kepadatan yang signifikan. Analisis statistik lebih lanjut melibatkan pengecekan outlier dan distribusi data. Selanjutnya, model regresi linier digunakan untuk memprediksi pertumbuhan populasi, dengan hasil evaluasi kinerja model menunjukkan MAPE sebesar 1860.78, yang dapat memberikan pandangan kritis terhadap akurasi model tersebut. Meskipun demikian, analisis ini memberikan wawasan mendalam tentang karakteristik pertumbuhan populasi dan faktor-faktor yang dapat mempengaruhinya.

V. UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih yang tulus kepada seluruh anggota kelompok atas kerja keras dan dedikasi kalian selama mengerjakan proyek ini. Kami juga ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada Bapak/Ibu Dosen Sains Data atas bimbingan dan materi yang telah diberikan selama proses penelitian ini. Semoga

penelitian ini dapat berkontribusi untuk masa depan yang lebih baik.

VI. DAFTAR PUSTAKA

1. “World Population Prospects - Population Division - United Nations.” n.d. Population Division |. Accessed December 29, 2023. <https://population.un.org/wpp/>.
2. “Analisis Data Eksploratif - Materi Lengkap dan Mendalam.” n.d. Smartstat. Accessed December 29, 2023. <https://www.smartstat.info/materi/statistika/eksplorasi-data.html>.
3. Ade, Setiawan. 2023. “Analisis Data Eksploratif: Teknik Aritmatika Sederhana dan Grafis untuk Meringkas Data.”
4. 2021. Anaconda | The World's Most Popular Data Science Platform. <https://www.anaconda.com/>.
5. “Mengenal Lebih Jauh Mengenai Anaconda Python - idmetafora.” 2022. software erp idmetafora. <https://idmetafora.com/news/read/1281/Mengenal-Lebih-Jauh-Mengenai-Anaconda-Python.html>.
6. D, Kuhlman. 2013. “A Python Book.” 1-227.
7. A.N, Syahrifudin and T. Kurniawan. 2018. “Input Dan Output Pada Bahasa.” 1-7.
8. Ade Setiawan. 2023. “Stemplot: Memahami dan Membuat Stem-and-Leaf Plot dalam Eksplorasi Data.” (Agustus).
9. Junaidi, J. 2015. “Deskripsi Data Melalui Box-Plot.” https://www.researchgate.net/publication/278022464_Deskripsi_Data_Melalui_Box-Plot.
10. Bologna, M, Aquino, G. n.d. “Deforestation and world population sustainability: a quantitative analysis.” Blog. <https://doi.org/10.1038/s41598-020-63657-6>.