



PROJECT ORACLE UAS

DTS-TSA March 2023 [Oracle Cloud Infrastructure]

Judul Penelitian : "Analisis Statistik dan Prediksi
Pertumbuhan Populasi Negara dengan Metode Regresi
Linear"



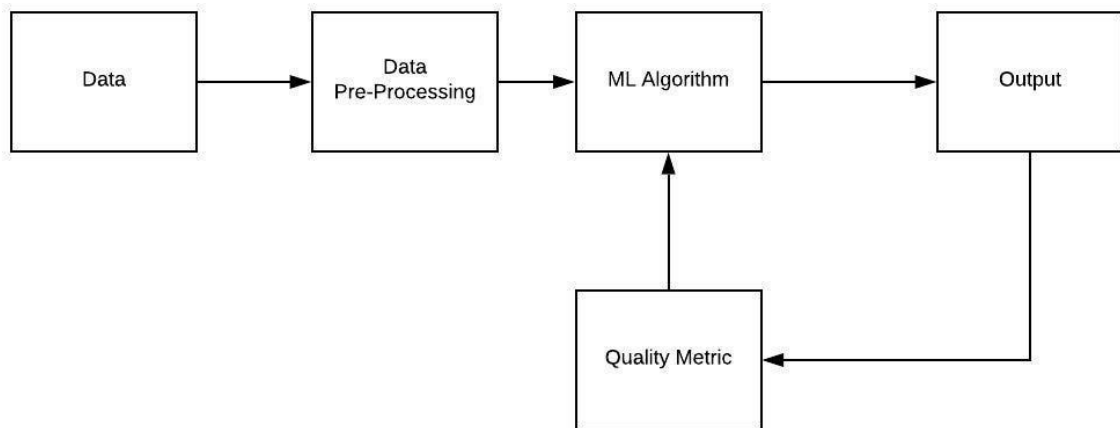
Anggota :

❖ **ALDOWAD ALLES SANDRO HAMONANGAN SIMANJUNTAK**

Talent Scouting Academy from Universitas Prima Indonesia



Flowchart Algoritma Linear Regresion



Langkah-langkah dalam regresi linear adalah sebagai berikut:

Persiapkan Data: Kumpulkan dan persiapkan dataset yang akan digunakan. Pastikan data sudah terstruktur dengan baik, bersih dari missing values, dan siap untuk dianalisis.

Tentukan Variabel: Identifikasi variabel independen (variabel penjelas) dan variabel dependen (variabel yang ingin diprediksi). Pastikan variabel tersebut memiliki hubungan linier yang dapat dijelaskan melalui regresi linear.

Pisahkan Data: Bagi dataset menjadi data pelatihan (training data) dan data uji (testing data). Data pelatihan digunakan untuk membangun model regresi linear, sedangkan data uji digunakan untuk menguji kinerja model yang telah dibangun.

Bangun Model: Gunakan fungsi atau library yang tersedia (seperti scikit-learn dalam Python) untuk membangun model regresi linear. Cocokkan model dengan data pelatihan menggunakan metode yang sesuai, seperti Metode OLS (Ordinary Least Squares) atau Metode Gradient Descent.

Evaluasi Model: Evaluasi kualitas model regresi linear yang telah dibangun dengan menggunakan metrik evaluasi yang relevan, seperti R-Squared (koefisien determinasi), Mean Squared Error (MSE), atau Mean Absolute Error (MAE). Nilai-nilai ini membantu mengevaluasi sejauh mana model cocok dengan data dan seberapa baik model dapat memprediksi.

Prediksi: Gunakan model yang telah dibangun untuk melakukan prediksi pada data uji atau data baru yang belum dilihat sebelumnya. Dapatkan nilai prediksi untuk variabel dependen berdasarkan nilai variabel independen yang diberikan.

Interpretasi Hasil: Interpretasikan hasil regresi linear, termasuk nilai koefisien regresi dan intercept. Analisis koefisien regresi membantu memahami sejauh mana setiap variabel independen mempengaruhi variabel dependen.

Kesimpulan dan Penyajian Hasil: Jelaskan temuan dan kesimpulan berdasarkan analisis regresi linear. Presentasikan hasil secara jelas dan sederhana, dan berikan rekomendasi atau implikasi berdasarkan hasil prediksi dan interpretasi model.



Kelebihan Oracle Machine Learning (OML):	Kelebihan Google Colab:
<p>Integrasi dengan Oracle Database: OML terintegrasi secara langsung dengan Oracle Database, memungkinkan analisis dan pemrosesan data langsung di dalam database tanpa perlu mentransfer data ke lingkungan lain. Ini mengurangi overhead dan meningkatkan efisiensi dalam penggunaan sumber daya.</p> <p>Performa Skala Besar: OML dapat mengambil keuntungan dari kemampuan komputasi yang kuat yang disediakan oleh Oracle Database, termasuk pemrosesan paralel dan distribusi otomatis yang dioptimalkan untuk kinerja skala besar. Ini memungkinkan pemrosesan cepat dan efisien dari volume data yang besar.</p> <p>Keamanan: OML dilengkapi dengan keamanan yang ketat yang terintegrasi dengan Oracle Database, termasuk kontrol akses, enkripsi, dan kepatuhan keamanan yang tinggi. Ini memberikan lapisan perlindungan tambahan bagi data dan model machine learning yang dihasilkan.</p> <p>Skalabilitas dan Fleksibilitas: OML dapat dengan mudah diukur sesuai kebutuhan dengan menggunakan sumber daya yang tersedia di Oracle Database. Anda dapat mengelola dan mengoptimalkan skala pemrosesan sesuai kebutuhan, memungkinkan penggunaan sumber daya yang efisien dan penyebaran model yang mudah.</p>	<p>Akses Mudah dan Gratis: Google Colab dapat diakses secara online melalui browser dan tersedia secara gratis. Ini memungkinkan pengguna untuk mengakses lingkungan pengembangan machine learning yang siap pakai tanpa perlu melakukan instalasi atau konfigurasi rumit.</p> <p>Ekosistem Google Cloud Services: Google Colab terintegrasi dengan layanan Google Cloud, seperti Google Drive dan BigQuery, memudahkan penggunaan dan analisis data dari sumber data yang berbeda.</p>

Pilihan antara Oracle Machine Learning dan Google Colab tergantung pada kebutuhan, preferensi, dan lingkungan kerja Anda. Jika Anda menggunakan Oracle Database dan membutuhkan integrasi langsung dengan database, OML dapat menjadi pilihan yang baik. Namun, jika Anda mencari platform yang mudah diakses dan memiliki kemampuan kolaborasi yang baik, Google Colab dapat menjadi pilihan yang cocok.

Dalam Project kali ini, peneliti menggunakan Oracle Machine Learning, berikut Langkah langkahnya :

LANGKAH 1 : MEMBUAT AUTONOMOUS DATABASE

Untuk mengolah data menggunakan Oracle Machine Learning AutoML UI di Oracle Cloud Autonomous Database dan mengimpor dataset World Population Data 1960-2020 dari komputer lokal, berikut adalah langkah-langkah:

Membuat Autonomous Database di Oracle Cloud Infrastructure:

Masuk ke Oracle Cloud Infrastructure Console (<https://console.cloud.oracle.com/>).

Pilih layanan Autonomous Database dan buat instance baru.

Ikuti langkah-langkah yang ditentukan dalam proses pembuatan untuk mengkonfigurasi detail seperti kompartemen, jenis database, kapasitas penyimpanan, dan kredensial administratif.

Create Autonomous Database

Provide basic information for the Autonomous Database

Compartment

aldowadsimanjuntak (root)

Display name

AutonomousDatabaseProject

A user-friendly name to help you easily identify the resource.

Database name

AutonomousDatabaseProject

The name must contain only letters and numbers, starting with a letter. Maximum of 30 characters.

Choose a workload type

Data Warehouse

Built for decision support and data warehouse workloads.

Transaction Processing

Built for transactional work-

JSON

Built for JSON-centric applica-
tion development. Developer-

APEX

Built for Oracle APEX applica-
tion development. Creation and

Configure the database

Always Free ⓘ

☐ Show only Always Free configuration options

Choose database version

19c

OCPU count

2

The number of OCPU cores to enable. Available cores are subject to your tenancy's service limits.

Storage (TB)

1

The amount of storage to allocate.

☐ OCPU auto scaling

Allows systems to expand up to three times the specified OCPU count as demand increases. [Learn more](#) about auto scaling.

☐ Storage auto scaling

Allows system to expand up to three times the reserved storage.

user here.

Create administrator credentials ⓘ

Username *Read-only*

ADMIN

ADMIN username cannot be edited.

Password

.....

Confirm password

.....

Choose license and Oracle Database edition

Choose a license type

Bring your own license (BYOL)

Bring your organization's Oracle Database software license to the Oracle Database service. [Learn more.](#)



License included

Subscribe to a new Oracle Database software license and the Oracle Database service.

Choose an Oracle Database Edition

Oracle Database Enterprise Edition (EE)

User-selected OCPU limit. [Learn more.](#)



Oracle Database Standard Edition (SE)

Up to 8 OCPUs, including auto scale. [Learn more.](#)

Provide contacts for operational notifications and announcements ⓘ

Contact Email

aldowadallessandrohamonangansimanjuntak@unprimdn.ac.id



LANGKAH 2 : KONFIGURASI DATABASE

Pilih Autonomous Database yang telah Anda buat sebelumnya.

Overview » Autonomous Database » Autonomous Databases

Autonomous Database

Autonomous Database

Dedicated infrastructure

Autonomous container database

Autonomous Exadata VM cluster

Exadata infrastructure

Create Autonomous Database

Display name	State	Dedicated	Compute	Storage	Workload type	Disaster recovery	Created
AutonomousDatabaseProject	Available	No	2	1 TB	Data Warehouse	Primary	Mon, May 29, 2023, 20:57:37 UTC

Displaying 1 Autonomous Database < 1 of 1 >

Download client credentials (Wallet) digunakan untuk mengunduh kredensial klien dalam bentuk wallet yang diperlukan untuk melakukan koneksi dengan menggunakan protokol mTLS (mutual Transport Layer Security) ke Autonomous Database.

Wallet ini mengamankan koneksi dan memastikan bahwa hanya klien yang memiliki sertifikat yang valid yang dapat terhubung ke database.

Overview » Autonomous Database » Autonomous Database details

AutonomousDatabaseProject

Database actions

Autonomous Database information

General information

Database name: AutonomousDatabaseProject

Workload type: Data Warehouse

Compartment: aldowadsimanjuntak (root)

OCID: ...przkyq Show Copy

Created: Mon, May 29, 2023, 20:57:37 UTC

OCPU count: 2

OCPU auto scaling: Disabled

Storage: 1 TB

Storage auto scaling: Disabled

License type: Bring your own license

Database version: 19c

Database connection

Connections to your Autonomous Database are secured, and can be authorized using TLS or mTLS authentication options. TLS authentication is easier to use, provides better connection latency, and does not require you to download client credentials (wallet) if any of these is true for your connections:

- You are using JDBC Thin Client (version 12.2.0.1 or higher) with JDK 8(u163+) or higher.
- You are using the Python python-oracledb driver.
- You are using ODP.NET version 19.14 (or higher), or 21.5 (or higher).
- You are using an Oracle Call Interface based driver with Oracle Client libraries version 19.14 (or higher), or 21.5 (or higher).

Learn more about TLS authentication and how to enable it.

Download client credentials (Wallet)

To download your client credentials, select the wallet type, and click **Download wallet**. You then enter a password for the wallet. This client credential download only contains information for mTLS connections. **You do not need a wallet for TLS connections.**

Wallet type

Instance Wallet

Download wallet Rotate wallet

Wallet last rotated: -

Close

Setelah itu Pilih opsi "Open Database Actions".

Di halaman Database Actions, klik "Load Data" untuk membuka fitur Load Data.

Your Oracle Cloud account has SGD 400.00 left in your Free Trial. When your trial is over, your account is limited to Always Free resources. Upgrade at any time.

ORACLE Cloud

machine learning

US East (Ashburn)

Overview » Autonomous Database » Autonomous Database details

AutonomousDatabaseProject • Primary

Database actions Database connection Performance hub Manage scaling More actions

Autonomous Database information Tool configuration Tags

General information

Database name: AutonomousDatabaseProject

Workload type: Data Warehouse

Compartment: aldowadsimanjuntak (root)

OCID: ...przkyq Show Copy

Created: Mon, May 29, 2023, 20:57:37 UTC

OCPU count: 2

OCPU auto scaling: Disabled

Infrastructure

Dedicated infrastructure: No

Disaster recovery

Role: Primary Switchover

Backup

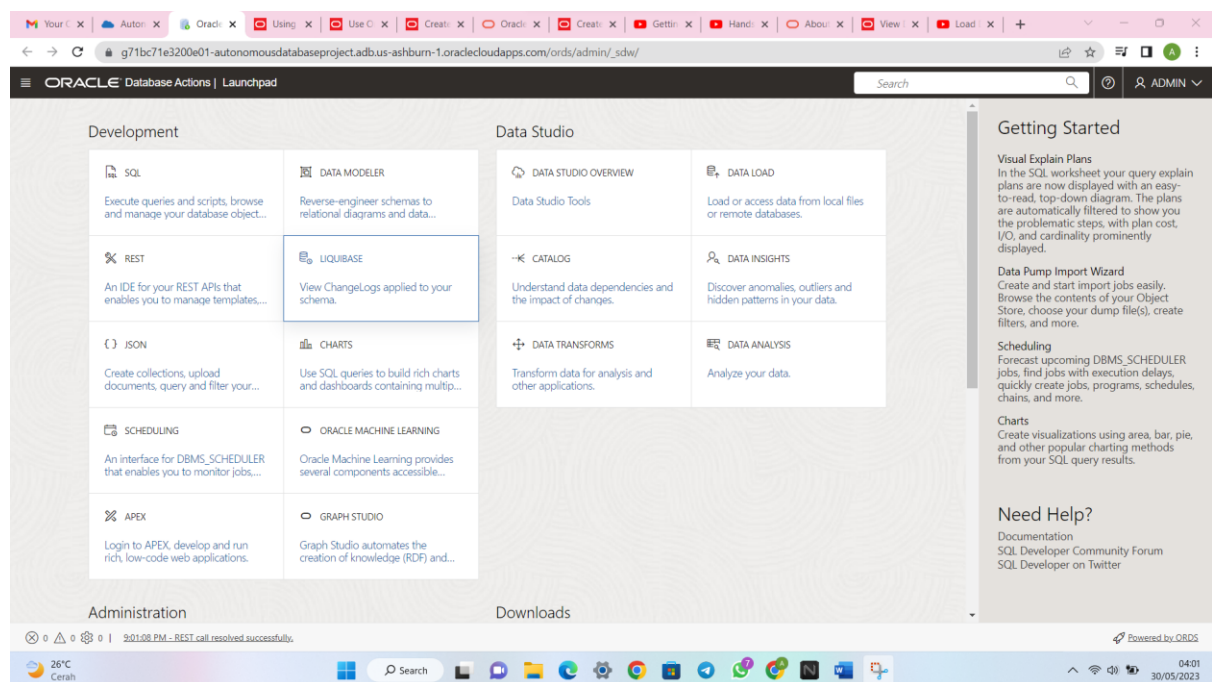
Announcements

- New Feature: Oracle APEX 23.1 is now available with new capabilities including template components, improved object browsing and push notifications. Learn more.
- Learn about other new features in Autonomous Database here.

Terms of Use and Privacy Cookie Preferences

Copyright © 2023. Oracle and/or its affiliates. All rights reserved.

Tampilan Interface Oracle Database Actions :



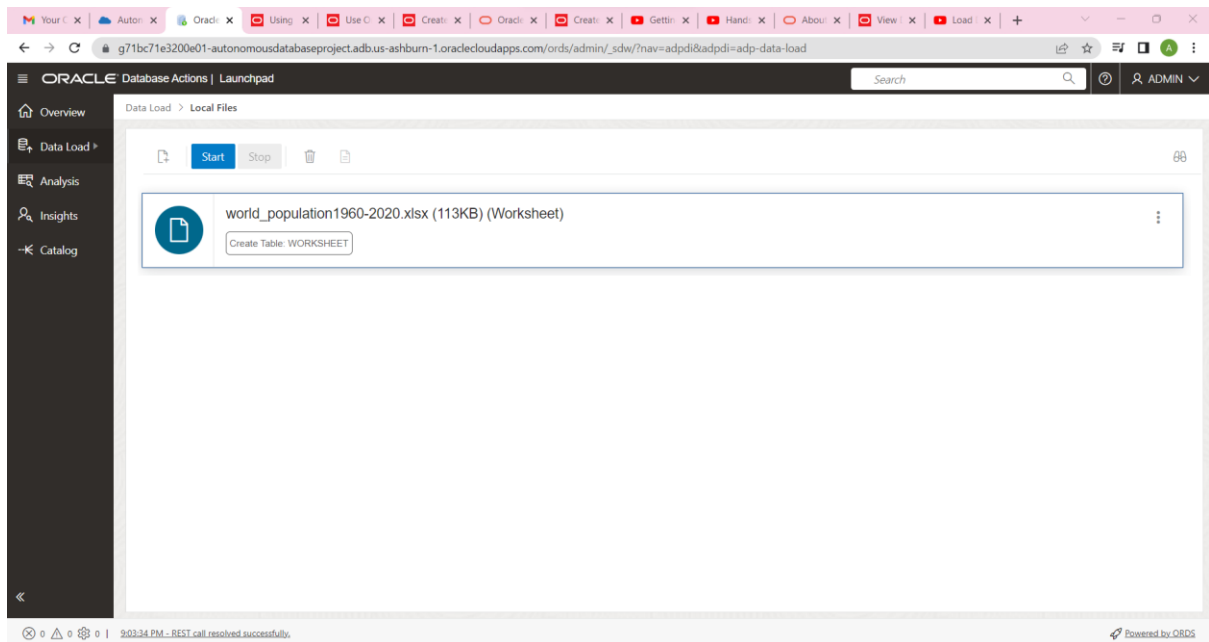
Mempersiapkan dataset: Pada komputer lokal kamu, pastikan dataset World Population Data 1960-2020 sudah siap dan berada dalam format yang sesuai. Dataset dapat berupa file CSV, Excel, atau format lain yang didukung oleh Oracle Autonomous Database.

Berikut adalah tampilan dataset di excel :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	country	year_1960	year_1961	year_1962	year_1963	year_1964	year_1965	year_1966	year_1967	year_1968	year_1969	year_1970	year_1971	year_1972	year_1973	year_1974	year_1975	year_1976	year_1977	year_1978	year_1979
2	Albania	8996967	9169406	9351442	9543200	9744772	9956318	10174840	10399936	10637064	10893772	11173654	11475450	11791222	12108963	12412960	12689164	12943093	13171294	13341199	1341110
3	Albania	1608800	1659800	1711319	1762621	1814135	1864791	1914573	1965598	2022272	2081695	2135479	2187853	2243126	2296752	2350124	2404831	2458526	2513546	2566266	26178
4	Algeria	11057864	11336336	11619828	11912800	12221675	12550880	12902626	13275020	13663581	14061724	14464992	14872253	15285992	15709831	16149018	16607706	17085799	17582899	18102266	186478
5	American Samoa	20127	20605	21246	22029	22850	23675	24473	25235	25980	26698	27362	27982	28564	29103	29595	30045	30455	30834	31262	318
6	Andorra	13410	14378	15379	16407	17466	18542	19646	20760	21886	23053	24275	25571	26885	28232	29515	30705	31782	32769	33744	348
7	Angola	5454938	5531451	5608499	5679409	5738995	5770573	5781305	5774440	5771973	5803677	5890360	6041239	6248965	6497283	6761623	7023994	7279630	7533814	7790774	80581
8	Antigua and Barbuda	54132	55005	55849	56701	57641	58699	59912	61240	62523	63553	64184	64354	64134	63649	63108	62671	62353	62162	62038	619
9	Argentina	20481781	20817270	21153042	21488916	21824427	22159644	22494031	22828872	23168268	23517613	23880564	24259564	24653172	25056475	25462305	25865775	26264681	26661397	27061041	2747110
10	Armenia	1874119	1941498	2009524	2077584	2145004	2211316	2276038	2339133	2401142	2462938	2525067	2587716	2650484	2712780	2773750	2832752	2889583	2944375	2997419	30491
11	Aruba	54208	55434	56234	56699	57029	57357	57702	58044	58377	58734	59070	59442	59849	60236	60527	60653	60586	60366	60102	599
12	Australia	10276477	10483000	10742000	10950000	11167000	11388000	11651000	11799000	12009000	12263000	12507000	12937000	13177000	13380000	13723000	13893000	14033000	14192000	14358000	1451400
13	Austria	7047539	7086299	7129864	7175811	7223801	7270889	7322066	7376998	7415403	7441055	7467086	7500482	7544201	7586115	7599038	7578903	7565525	7568430	7562305	75494
14	Azerbaijan	3895398	4030325	4171428	4315127	4456691	4592601	4721528	4843872	4960237	5071930	5180032	5284518	5385266	5483088	5579071	5674129	5768730	5863138	5957927	60536
15	Bahamas	109532	115119	121092	127340	133705	140060	146381	152621	158648	164265	169376	173894	177863	181519	185105	188895	192903	197100	201482	20597
16	Bahrain	162429	167899	173140	178142	182888	187432	191785	196060	200652	206037	212607	220311	229151	239527	251908	266540	283746	303169	323468	3428
17	Bangladesh	48013505	49362834	50752150	52202008	53741721	55385114	57157651	59034250	60918452	62679765	64232486	65531635	66625706	67637541	68742222	70066310	71652386	73463593	75450033	775290
18	Barbados	230985	231718	232623	233632	234588	235415	236084	236661	237241	237963	238895	240093	241523	243076	244643	246158	247584	248931	250200	2513
19	Belarus	8198000	8271216	8351928	8437232	8524224	8610000	8696496	8785648	8874552	8960304	9040000	9115576	9188968	9257272	9317584	9367000	9411000	9463000	9525000	95840
20	Belgium	9153489	9183948	9220578	9289770	9378113	9463667	9527807	9580991	9618756	9646032	9655549	9673162	9711115	9741720	9772419	9800700	9818227	9830358	9839534	98483
21	Belize	92068	94700	97392	100165	103069	106120	109348	112707	116065	119269	122184	124795	127152	129294	131305	133264	135143	136990	138975	1413
22	Benin	2431617	2465865	2502897	2542864	2585961	2632361	2682159	2735308	2791588	2850657	2912338	2976575	3043563	3113681	3187413	3265167	3347169	3432445	3523933	36185
23	Bermuda	44400	45500	46600	47700	48900	50100	51000	52000	53000	54000	55000	56000	57000	58000	59000	60000	61000	62000	63000	64000
24	Bhutan	223284	228849	234552	240529	246961	253993	261664	269994	278731	287886	297307	306957	316822	326986	337491	348395	359721	371424	383234	3951
25	Bolivia	3656961	3728954	3802996	3879191	3957759	4038865	4122517	4208883	4297522	4389248	4484004	4581752	4682392	4785916	4892294	5001413	5113458	5228253	5344946	54624
26	Bosnia and Herzegovina	2386622	2438662	2493333	2550666	2610666	2673333	2738666	2806666	2877666	2951666	3028666	3108666	3191666	3277666	3368666	3462666	3559666	3659666	3762666	3869666

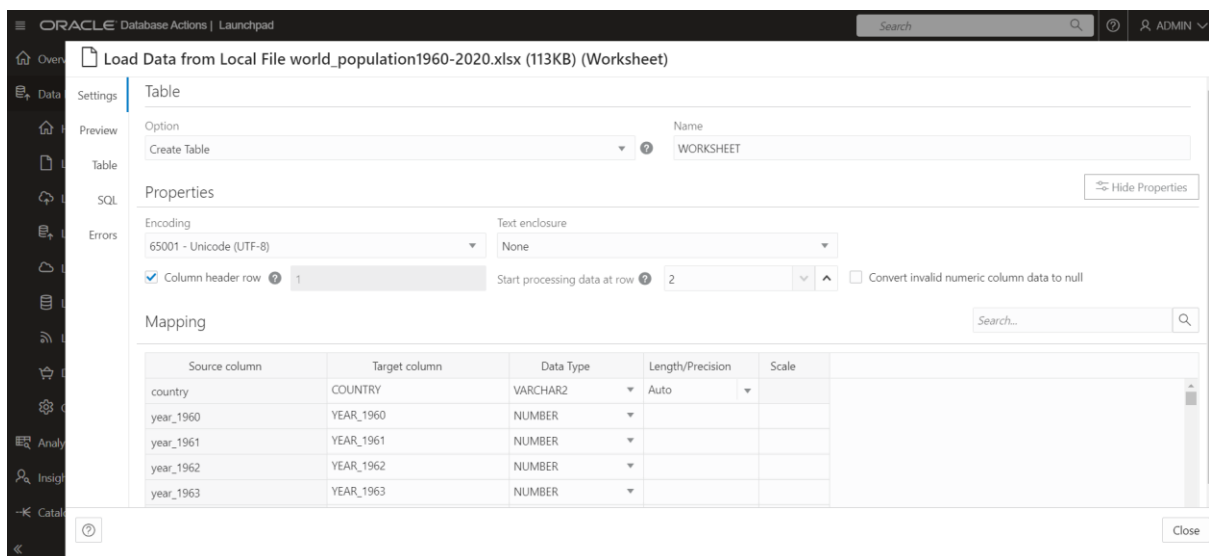
LANGKAH 3 : LOAD DATASET KE ORACLE AUTONOMOUS DATABASE

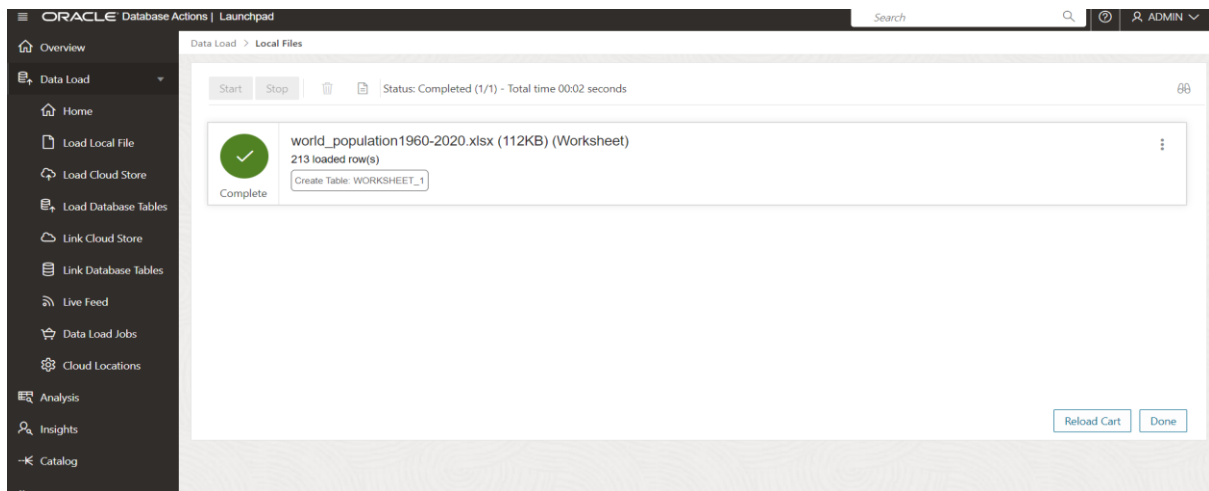
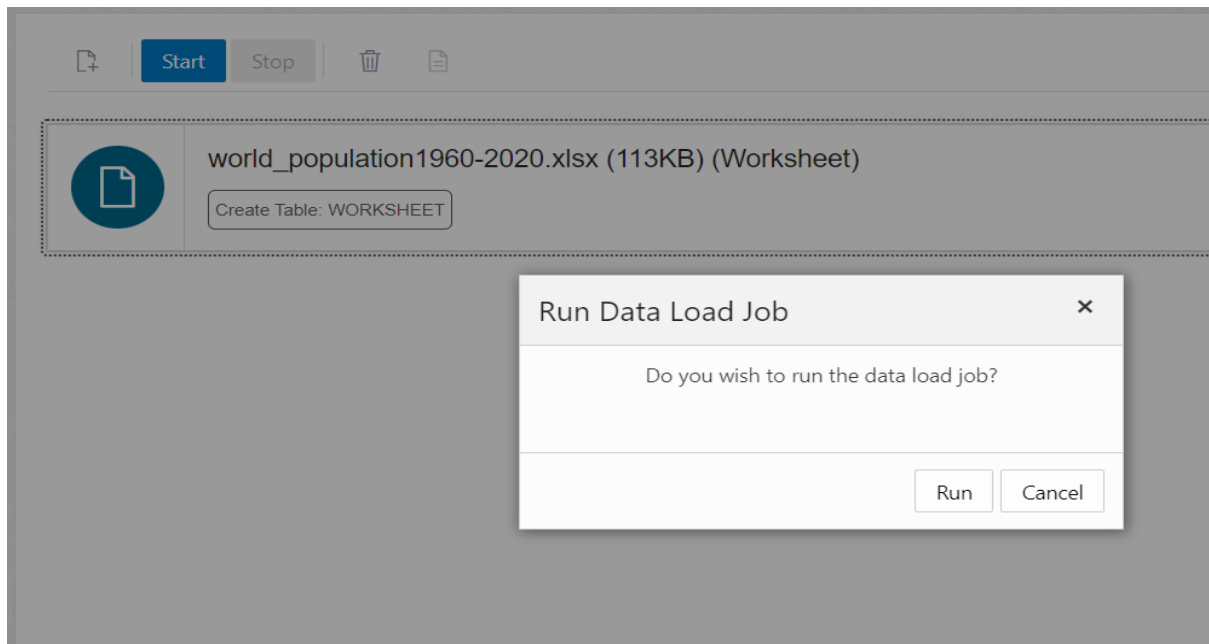
Impor dataset ke Autonomous Database: Gunakan alat seperti SQL Developer atau perintah SQL untuk mengimpor dataset dari komputer lokal ke Autonomous Database. Anda bisa menggunakan perintah SQL LOAD DATA atau IMPORT DATA untuk melakukan impor dataset ke dalam tabel yang sudah dibuat di database.



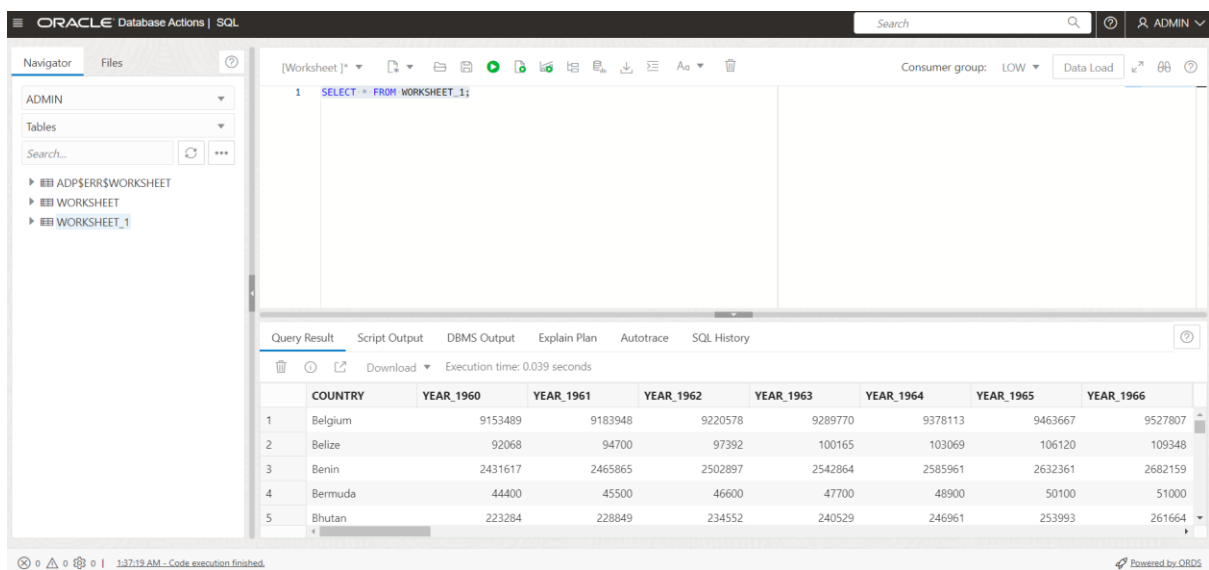
Klik tombol "Load Data" untuk memulai proses impor dataset ke tabel yang ditentukan.

Tunggu hingga proses import selesai. Anda akan melihat status impor dan riwayat impor di halaman Database Actions.



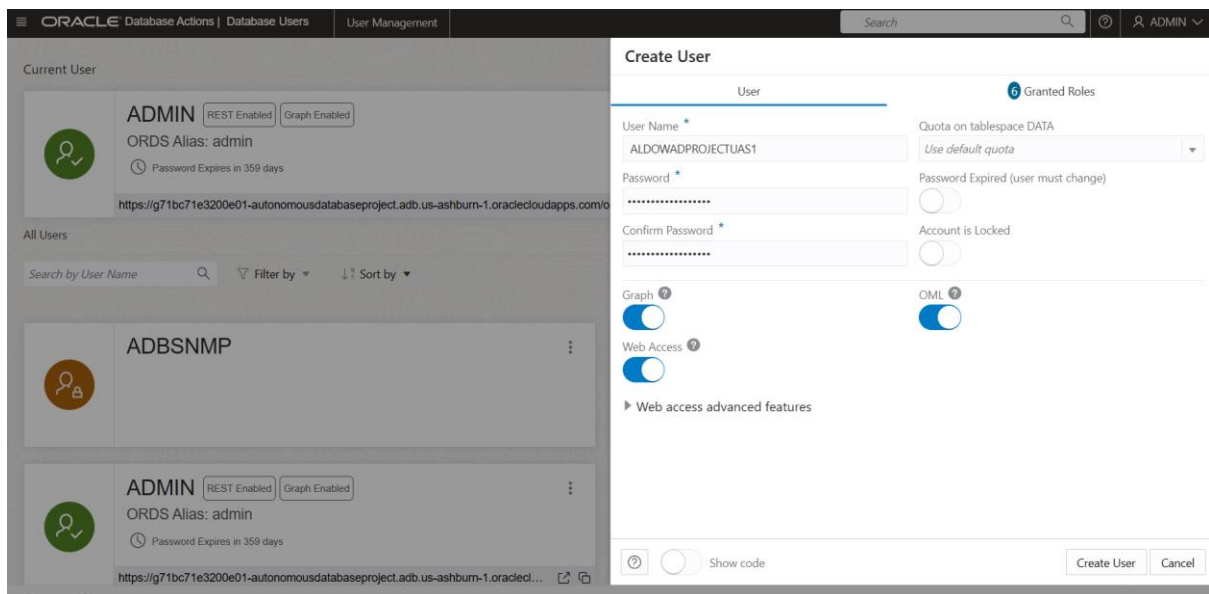


Buka Developer SQL Tools untuk mengecek apakah dataset bisa dipanggil dengan SQL :

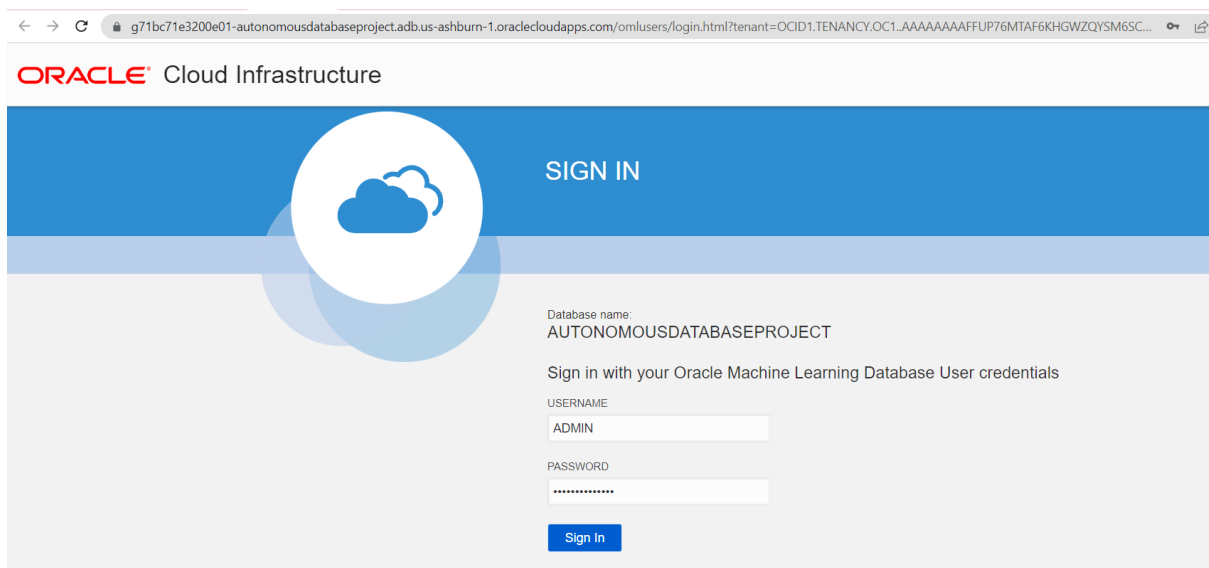


LANGKAH 4 : BUAT AKUN ORACLE MACHINE LEARNING

Sebelum masuk ke Oracle Machine Learning, buat User Credential agar Ketika Membuka Notebook datasetnya sudah terkoneksi :



Login menggunakan akun yang sudah dibuat tadi :



Tampilan Interface Oracle Machine Learning :

The screenshot displays the Oracle Machine Learning (OML) interface. The top navigation bar includes the 'ORACLE Machine Learning' logo and the 'ADMIN Project' dropdown. Below the navigation bar, there are three main sections: 'How Do I?', 'Quick Actions', and 'Recent Activities'. The 'How Do I?' section contains links for 'Get Started', 'Use AutoML', 'Deploy Models', 'Create Notebooks', and 'Create Jobs'. The 'Quick Actions' section contains links for 'AutoML', 'Models', 'Scratchpad', 'Notebooks', 'Jobs', and 'Examples'. The 'Recent Activities' section shows 'Nothing to Display'.

LANGKAH 5 : VISUALISASI DATA DENGAN SQL

Buat Notebook baru

The screenshot shows the 'Notebooks' section of the Oracle Machine Learning interface. At the top, there is a 'Go to OML Notebooks EA' button. Below it, there is a toolbar with various actions: '+ Create', 'Edit', 'Delete', 'Duplicate', 'Move', 'Copy', 'Save as Template', 'Import', 'Export', 'Version', and 'Copy to OML Notebooks EA'. A 'Filter' input field is also present. The main table lists the notebooks, with one notebook named 'Project_Oracle' shown. The table columns are: Name, Comment, Created On, Created By, Last Update, and Updated By.

Name	Comment	Created On	Created By	Last Update	Updated By
Project_Oracle		5/30/23, 9:39 AM	ADMIN	6/3/23, 12:04 AM	ADMIN

Setelah dibuka, mulai jalankan perintah SQL :

SELECT * FROM WORKSHEET_1: Untuk mengambil semua data (semua kolom dan semua baris) dari tabel "WORKSHEET_1" kedalam Notebook Oracle Machine Learning

The screenshot shows the 'Project_Oracle' notebook in the Oracle Machine Learning interface. The notebook is titled 'Project_Oracle' and contains a SQL query: 'SELECT * FROM WORKSHEET_1;'. The query has been executed, and the results are displayed in a table. The table has columns: COUNTRY, YEAR_1960, YEAR_1961, YEAR_1962, YEAR_1963, YEAR_1964, YEAR_1965, YEAR_1966, YEAR_1967, and YEAR_1968. The data is sorted by COUNTRY.

COUNTRY	YEAR_1960	YEAR_1961	YEAR_1962	YEAR_1963	YEAR_1964	YEAR_1965	YEAR_1966	YEAR_1967	YEAR_1968
Belgium	9153489	9183948	9220578	9289770	9378113	9463667	9527807	9580991	9618113
Belize	92068	94700	97392	100165	103069	106120	109348	112707	1160113
Benin	2431617	2465865	2502897	2542864	2585961	2632361	2682159	2735308	2791113
Bermuda	44400	45500	46600	47700	48900	50100	51000	52000	53000
Bhutan	223284	228849	234552	240529	246961	253993	261664	269944	2787113
Bolivia	3656961	3728954	3802996	3879191	3957759	4038865	4122517	4208683	4297113
Bosnia and Herzegovina	3225664	3288604	3353228	3417573	3478999	3535632	3586630	3632678	3677113

Took 13 secs. Last updated by ADMIN at June 02 2023, 11:44:11 PM. (outdated)

Memeriksa tipe data dari semua kolom dalam table :

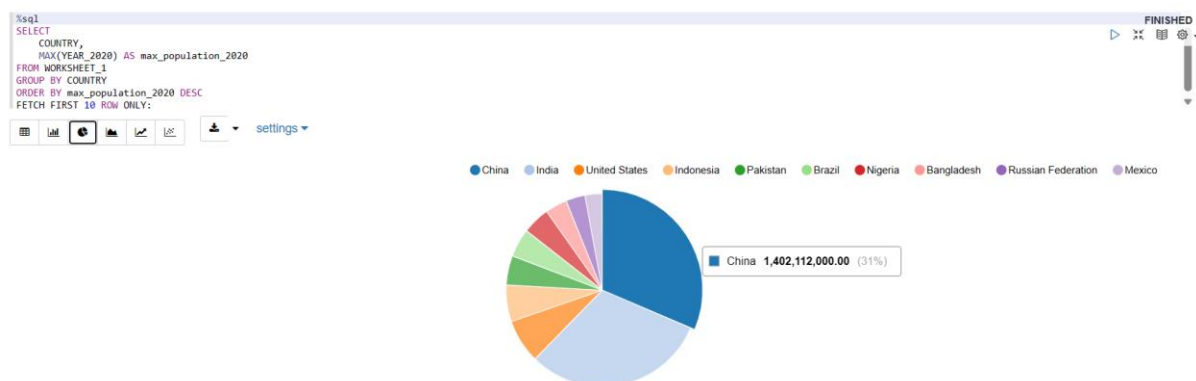
```
%sql
SELECT COLUMN_NAME, DATA_TYPE
FROM ALL_TAB_COLUMNS
WHERE TABLE_NAME = 'WORKSHEET_1';
```

COLUMN_NAME	DATA_TYPE
COUNTRY	VARCHAR2
YEAR_1960	NUMBER
YEAR_1961	NUMBER
YEAR_1962	NUMBER
YEAR_1963	NUMBER
YEAR_1964	NUMBER
YEAR_1965	NUMBER
YEAR_1966	NUMBER

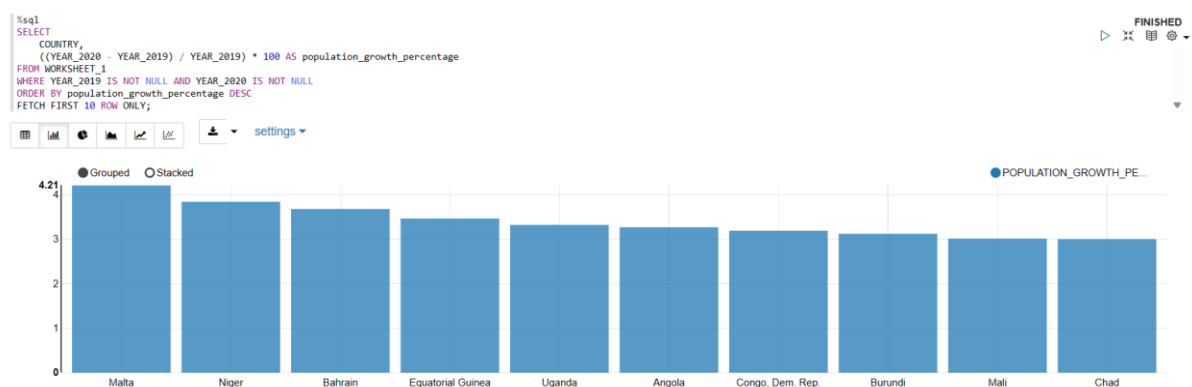
Pengolahan Data :

EDA (Exploratory Data Analysis) adalah proses eksplorasi dan analisis awal yang dilakukan pada data sebelum menerapkan model atau menjalankan penelitian lebih lanjut. Tujuan utama EDA adalah untuk memahami karakteristik data, mengidentifikasi pola, mengungkap hubungan antara variabel, dan mendapatkan wawasan yang berguna dari data.

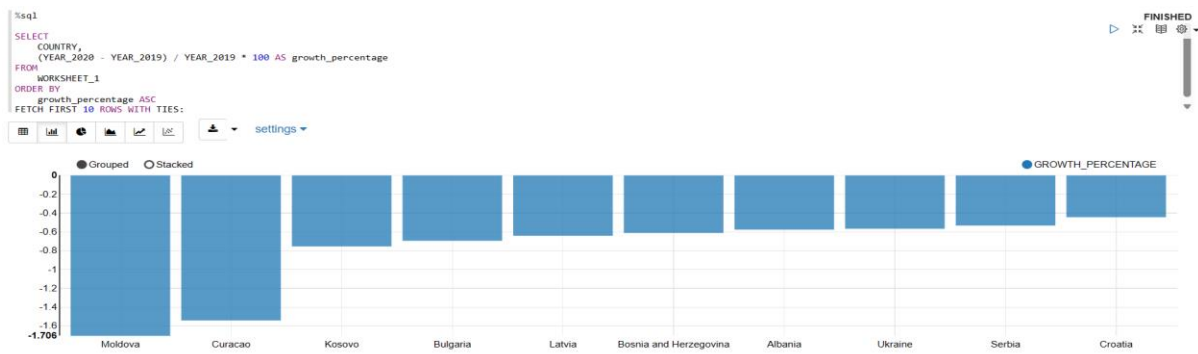
Identifikasi 10 negara dengan pertumbuhan populasi tertinggi negara pada tahun 2020:



Persentase pertumbuhan penduduk tertinggi dari tahun 2019 ke tahun 2020



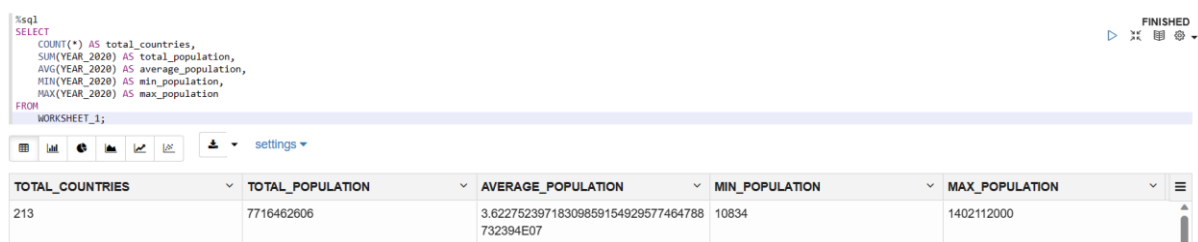
Persentase pertumbuhan penduduk terendah dari tahun 2019 ke tahun 2020



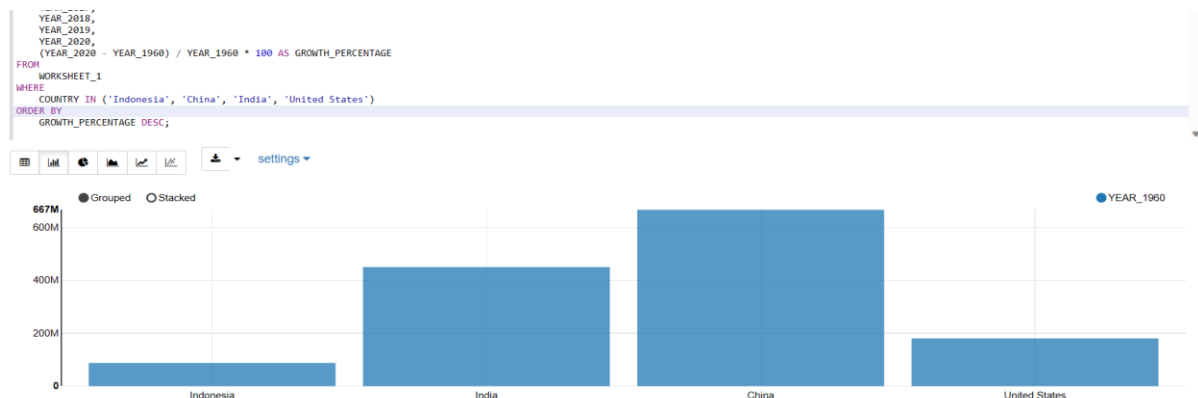
Menampilkan pertumbuhan negara China, India, United States, dan Indonesia antara tahun 1960 dibandingkan tahun 2020:



Statistik Deskriptif: menggunakan fungsi agregasi seperti COUNT, SUM, AVG, MIN, MAX, dll. untuk menghitung total penduduk, rata-rata penduduk, penduduk maksimum, penduduk minimum :



Persentasi Pertumbuhan Penduduk dari tahun 1960 sampai 2020 negara Indonesia, China, India dan United States :



LANGKAH 6 : KONEKSI DATABASE

Untuk membuat Koneksi database tadi agar bisa diolah dengan python di Oracle Machine Learning

Kembali ke Oracle untuk mendapatkan user, password dan DSN

Overview » Autonomous Database » Autonomous Database details

Autonomous Database

Database actions

Autonomous Database info

General information

Database name:

Autonomous Database

Workload type:

Data Warehouse

Compartment:

aldowadsima

OCID:

...prklyq [Show](#) [Copy](#)

Created:

Mon, May 29, 2023

OCPU count:

2

OCPU auto scaling:

Disable

Storage:

1 TB

Storage auto scaling:

Disable

License type:

Bring your own

Database connection

Help

Wallet last rotated: -

Connection Strings

Use the following connection strings or TNS names for your connections. See the [documentation](#) for details.

TLS Authentication

Mutual TLS

TNS Name ⓘ	Connection String ⓘ
autonomousdatabaseproject_high	...ecurity=(ssl_server_dn_match=yes))) Show Copy
autonomousdatabaseproject_low	...ecurity=(ssl_server_dn_match=yes))) Show Copy
autonomousdatabaseproject_medium	...ecurity=(ssl_server_dn_match=yes))) Show Copy

Showing 3 items

Close

Mencoba koneksi untuk memanggil dataset tadi dengan program Python :

```
%python
import cx_Oracle

# Membuat koneksi ke database
connection = cx_Oracle.connect(
    user="ADMIN",
    password="Pr0ject@d@w",
    dsn="(description=(retry_count=20)(retry_delay=3)(address=(protocol=tcp)(port=1522)(host=adb.us-ashburn-1.oraclecloud.com))(connect_data=(service_name=g71bc71e3200e01_autonomousdatabaseproject_high
    .adb.oraclecloud.com))(security=(ssl_server_dn_match=yes)))"
)

# Membuat cursor
cursor = connection.cursor()

# Menjalankan query untuk mengambil data dari tabel
query = "SELECT * FROM WORKSHEET_1"
cursor.execute(query)

# Membaca hasil query
for row in cursor:
    # Lakukan sesuatu dengan data di setiap baris
    print(row)

# Menutup cursor dan koneksi
cursor.close()
connection.close()

('Cayman Islands', 7870, 8024, 8141, 8219, 8299, 8370, 8444, 8518, 8638, 8833, 9143, 9588, 10135, 10778, 11488, 12234, 13023, 13848, 14681, 15485, 16207, 16857, 17436, 18005, 18662, 19460, 20426, 21539, 2277
5, 24028, 25307, 26540, 27784, 29068, 30521, 32161, 34065, 36155, 38330, 40423, 42305, 43934, 45347, 46624, 47899, 49261, 50729, 52280, 53835, 55321, 56672, 57877, 58963, 59933, 60848, 61721, 62564, 63382, 6
4172, 64948, 65720)
('Central African Republic', 1501668, 1526057, 1551908, 1579375, 1608618, 1639706, 1673019, 1708306, 1744198, 1778870, 1811157, 1840517, 1867786, 1894850, 1924386, 1958367, 1997017, 2039914, 2087662, 214077
8, 2199359, 2264441, 2335339, 2408322, 2478382, 2542170, 2597765, 2646836, 2693974, 2745735, 2806740, 2878507, 2959236, 3046148, 3135017, 3222662, 3308235, 3392432, 3475485, 3558019, 3640421, 3722016, 380212
9, 3881185, 3959883, 4038380, 4118075, 4198004, 4273368, 4337623, 4386765, 4418639, 4436411, 4447945, 4464171, 4493171, 4537683, 4596023, 4666375, 4745179, 4829764)
('Chad', 3001604, 3060365, 3121226, 3183576, 3246527, 3309583, 3372182, 3434817, 3499370, 3568402, 3643608, 3726189, 3815253, 3907891, 3999918, 4088568, 4173131, 4255242, 4337292, 4422743, 4514427, 4612858,
4718157, 4832316, 4957561, 5095400, 5247281, 5412844, 5589624, 5773930, 5963250, 6157085, 6356741, 6563925, 6781057, 7010159, 7250974, 7503494, 7770053, 8053532, 8355654, 8678049, 9019226, 9373913, 9734761,
10096630, 10457122, 10818031, 11183589, 11560142, 11952134, 12360986, 12784748, 13220433, 13663562, 14110971, 14561658, 15016761, 15477727, 15946882, 16425859)
('Chile', 8132988, 8303804, 8476895, 8650390, 8821855, 8989607, 9152848, 9312091, 9468851, 9625304, 9783134, 9942716, 10103675, 10265827, 10428803, 10592310, 10756876, 10922777, 11089165, 11254877, 11419350,
11582020, 11743909, 11907955, 12078137, 12257238, 12445833, 12642917, 12847712, 13058758, 13274617, 13495255, 13719818, 13944934, 14166346, 14388064, 14587367, 14786227, 14977736, 15162081, 15342358, 1551611
2, 15684413, 15849409, 16014972, 16182713, 16354507, 16530201, 16708825, 16886184, 17062531, 17233584, 17408359, 17571511, 17738909, 17909356, 18209072, 18470435, 18729166, 18952035, 19116209)
('China', 667070000, 660330000, 665770000, 682335000, 698355000, 715185000, 735400000, 754550000, 774510000, 796025000, 818315000, 841105000, 862030000, 881940000, 900350000, 916395000, 930685000, 943455000,
956165000, 969005000, 981235000, 993885000, 1008630000, 1023310000, 1036825000, 1051040000, 1066790000, 1084035000, 1101630000, 1118650000, 1135185000, 1150780000, 1164970000, 1178440000, 1191835000, 1204855
000, 1217550000, 1230075000, 1241935000, 1252735000, 1262645000, 1271850000, 1280400000, 1288400000, 1296075000, 1303720000, 1311020000, 1317850000, 1324655000, 1331260000, 1337705000, 1344130000, 135069500
0, 1357380000, 1364270000, 1371220000, 1378665000, 1386395000, 1392730000, 1397715000, 1402112000)
('Colombia', 16057714, 16567817, 17092919, 17629978, 18175187, 18725242, 19279734, 19837508, 20393704, 20942453, 21480064, 22003983, 22516429, 23024512, 23538390, 24065502, 24608102, 25164544, 25733669, 2631
```

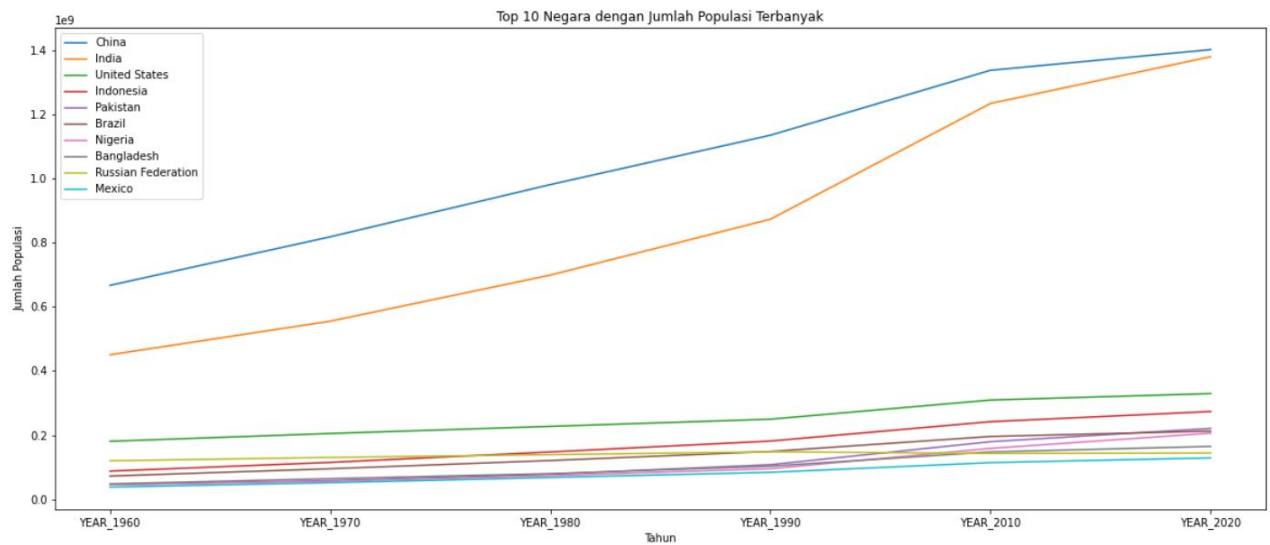
Menampilkan dataset :

%python
df

	COUNTRY	YEAR_1960	...	YEAR_2019	YEAR_2020
0	Cayman Islands	7870	...	64948	65720
1	Central African Republic	1501668	...	4745179	4829764
2	Chad	3001604	...	15946882	16425859
3	Chile	8132988	...	18952035	19116209
4	China	667070000	...	1397715000	1402112000
...
208	Marshall Islands	14674	...	58791	59194
209	Mauritania	850377	...	4525698	4649660
210	Mauritius	659351	...	1265711	1265740
211	Mexico	37771861	...	127575529	128932753
212	Micronesia, Fed. Sts.	44510	...	113811	115021

[213 rows x 62 columns]

Memvisualisasikan dataset kedalam Grafik :



LANGKAH 7 : PENGOLAHAN DATA DENGAN REGGRESION LINEAR

Persiapan data yang akan diolah :

```
%python
# Mengambil kolom negara dan kolom tahun-tahun yang diinginkan
selected_columns = ['COUNTRY'] + [f'YEAR_{year}' for year in range(1960, 2021)]
selected_data = data[selected_columns]
```

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:05 PM.

```
%python
# Memilih data untuk negara yang dipilih
countries = ['China', 'India', 'United States', 'Indonesia']
selected_data = selected_data[selected_data['COUNTRY'].isin(countries)]
```

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:05 PM.

```
%python
print(selected_data)
```

	COUNTRY	YEAR_1960	YEAR_1961	...	YEAR_2018	YEAR_2019	YEAR_2020
4	China	667070000	660330000	...	1392730000	1397715000	1402112000
24	United States	180671000	183691000	...	326838199	328329953	329484123
156	India	450547675	459642166	...	1352642283	1366417756	1380004385
157	Indonesia	87751066	90098396	...	267670549	270625567	273523621

[4 rows x 62 columns]

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:06 PM.

Membangun Model:

Import kelas LinearRegression dari Scikit-learn.

Buat objek model regresi linear menggunakan LinearRegression().

Latih model dengan menggunakan metode fit() pada data fitur (X) dan target (y).

```
%python
# Memisahkan data menjadi fitur (X) dan target (y)
X = selected_data.iloc[:, 1:].values # Mengambil semua kolom kecuali kolom pertama (negara)
y = selected_data['YEAR 2020'].values
```

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:06 PM.

```
%python
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
# Membangun model Regresi Linear
model = LinearRegression()
model.fit(X, y)
```

LinearRegression()

Evaluasi Model:

```
%python
# Evaluasi model
r_squared = model.score(X, y)
print('R-squared:', r_squared)
```

R-squared: 1.0

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:06 PM.

```
%python
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Memprediksi target menggunakan fitur X
y_pred = model.predict(X)

# Menghitung Mean Squared Error (MSE)
mse = mean_squared_error(y, y_pred)
print('Mean Squared Error (MSE):', mse)

# Menghitung Mean Absolute Error (MAE)
mae = mean_absolute_error(y, y_pred)
print('Mean Absolute Error (MAE):', mae)
```

Mean Squared Error (MSE): 4.440892098500626e-15
Mean Absolute Error (MAE): 4.470348358154297e-08

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:07 PM.

R-squared (Koefisien Determinasi): R-squared mengukur sejauh mana variasi dalam data target dapat dijelaskan oleh model. Rentang nilai R-squared adalah antara 0 hingga 1, dan semakin dekat ke 1, semakin baik modelnya. Dalam Penelitian kali ini didapatkan nilai 1, yang menandakan model dapat dengan baik memprediksi jumlah penduduk berdasarkan dataset sebelumnya.

Mean Absolute Error (MAE): MAE mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MAE, semakin baik performa modelnya. Anda dapat menggunakan metode `mean_absolute_error(y_true, y_pred)` dari modul `sklearn.metrics` untuk menghitung MAE.

Mean Squared Error (MSE): MSE mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MSE, semakin baik performa modelnya. Anda dapat menggunakan metode `mean_squared_error(y_true, y_pred)` dari modul `sklearn.metrics` untuk menghitung MSE.

Berikut Nilai MSE dan MAE dari penelitian ini :

Mean Squared Error (MSE): 4.44089

Mean Absolute Error (MAE): 4.47034

Dalam penelitian kali ini MSE memiliki nilai sangat kecil, mendekati nol. Ini menunjukkan bahwa kesalahan antara prediksi model dan nilai sebenarnya sangat kecil. MAE juga memiliki nilai yang sangat kecil, menunjukkan bahwa rata-rata kesalahan absolut antara prediksi model dan nilai sebenarnya juga sangat kecil.

Hal ini mengindikasikan bahwa model regresi linear kami memiliki tingkat akurasi yang tinggi dalam memprediksi populasi. Namun, tetap perlu diperhatikan bahwa hasil evaluasi ini didasarkan pada data yang digunakan untuk melatih model, dan kinerja model dapat berbeda ketika diterapkan pada data yang baru atau di luar rentang data pelatihan.

Prediksi untuk tahun 2021- 2025 dari model yang sudah dilatih :

```
%python
import pandas as pd

# Menentukan tahun-tahun yang ingin diprediksi
tahun_prediksi = [2021, 2022, 2023, 2024, 2025]

# Membuat DataFrame kosong untuk menyimpan hasil prediksi
hasil_prediksi = pd.DataFrame(columns=['Negara'] + [f'Tahun {tahun}' for tahun in tahun_prediksi])

# Melakukan prediksi untuk setiap tahun dan negara yang dipilih
for country in countries:
    idx = selected_data[selected_data['COUNTRY'] == country].index.item()
    populasi_sekarang = selected_data.loc[idx, 'YEAR_2020']

    prediksi_per_tahun = []
    for i, tahun in enumerate(tahun_prediksi):
        X_prediksi = selected_data.loc[idx, selected_columns[1:]].values.reshape(1, -1)
        X_prediksi[0, -1] = tahun

        prediksi = model.predict(X_prediksi)[0]
        persentase_tambahan = 0.1 + (i+2) * 0.01 # Menghitung persentase tambahan berdasarkan tahun
        prediksi = prediksi * (1 + persentase_tambahan)
        prediksi_per_tahun.append(int(prediksi))

    hasil_prediksi = hasil_prediksi.append({'Negara': country, **{f'Tahun {tahun}': prediksi for tahun in zip(tahun_prediksi, prediksi_per_tahun)}}, ignore_index=True)

# Menampilkan hasil prediksi dalam bentuk tabel
print(hasil_prediksi)
```

	Negara	Tahun 2021	Tahun 2022	Tahun 2023	Tahun 2024	Tahun 2025
0	China	1427161526	1439904040	1452646554	1465389067	1478131581
1	India	1404658949	1417200547	1429742145	1442283743	1454825341
2	United States	335370701	338365083	341359464	344353845	347348227
3	Indonesia	278410441	280896249	283382057	285867864	288353672

Menutup Koneksi ke Database Ketika sudah menyelesaikan Penelitian :

```
%python
connection.close()
```

Took 0 secs. Last updated by ADMIN at June 04 2023, 8:25:07 PM.

KESIMPULAN :

Berdasarkan hasil evaluasi ini, kami menyimpulkan bahwa model regresi linear yang kami latih dapat digunakan dengan percaya diri untuk memprediksi populasi negara pada tahun-tahun berikutnya. Namun, perlu diingat bahwa prediksi model didasarkan pada data historis dan kinerja model dapat bervariasi ketika diterapkan pada data baru di luar rentang pelatihan.

Penelitian ini memberikan kontribusi dalam memahami dan memprediksi pola pertumbuhan populasi negara dengan menggunakan pendekatan regresi linear. Hasilnya dapat menjadi dasar untuk pengambilan keputusan dalam berbagai konteks, seperti perencanaan pembangunan, kebijakan sosial, dan studi demografi.

Namun, perlu diingat bahwa setiap penelitian memiliki batasan dan asumsi tertentu. Dalam penelitian ini, kita mengasumsikan bahwa tren pertumbuhan populasi akan tetap konsisten di masa mendatang, yang mungkin tidak selalu terjadi dalam realitas. Selain itu, perubahan sosial, ekonomi, dan kebijakan juga dapat mempengaruhi pola pertumbuhan populasi di masa depan.

Oleh karena itu, untuk penelitian lanjutan, disarankan untuk melibatkan faktor-faktor lain seperti variabel ekonomi, sosial, dan demografi yang dapat mempengaruhi pertumbuhan populasi. Juga, melakukan validasi dan pengujian model pada data independen yang tidak digunakan dalam pelatihan dapat memberikan pemahaman yang lebih baik tentang kemampuan prediksi model.

Dengan demikian, penelitian ini menyediakan fondasi yang kuat untuk analisis populasi lebih lanjut dan memberikan wawasan yang berharga bagi pengambilan keputusan dan perencanaan di berbagai bidang yang terkait dengan populasi negara.