ENERGY CONSUMPTION OF NETHERLANDS

SPARK INDIVIDUAL ASSIGNMENT

SUBMITTED TO:

PROFESSOR RAUL MARIN PEREZ

SUBMITTED BY:

MANVI CHAUDHARY

TABLE OF CONTENTS:

Page No.

Dataset Link : https://www.kaggle.com/lucabasa/dutch-energy

INTRODUCTION

Energy, the core of our survival and development, helps us understand how a nation consumes its energy and what trends does it follow in consumption is crucially vital for future planning for utility providers.

Illuminating on the electricity consumption in Netherlands for past 10 years, from 2010 to 2019, Enexis, Liander, and Stedin(3 major administration firms) provide the energy to the entire country.

Year	Population	Yearly % Change
2019	17,097,130	0.22 %
2018	17,059,560	0.22 %
2017	17,021,347	0.24 %
2016	16,981,295	0.25 %
2015	16,938,499	0.30 %
2010	16,682,917	0.38 %

Yearly population growth(Netherlands)

Therefore, here i will try to analyze the electricity consumption in Netherlands.

GOAL OF ANALYSIS

Enexis, Liander, and Stedin every year release on their table summarizing the energy consumption of the areas under their administration, where they roughly succeed in equipping energy to the designated zip codes. Small deviations from 1 year to the other can be attributed to a change of management or for a different aggregation of zip codes.

With the aim to analyze the amount of electricity consumed individually per connection, a comparison between consumption and delivery percentage in certain areas serves the best solution with assigning responsibilities to the concerned network providers.

Smart meters spread across the cities assist in analyzing the cities with the highest number of smart meters installed and highlighting their respective journeys to reach this level of smart meters installation.

Areas segregated by streets assist in providing the top streets that receive the highest percentage of electricity.

Enexis, Liander, and Stedin analyze which cities receives the minimum or maximum amount of electricity and come up with a relevant conclusion whether it is related to the number of electricity connections present?

ANALYSIS DEEP DIVE

Dataset : https://www.kaggle.com/lucabasa/dutch-energy

QUESTIONS

1. What the consumption of electricity per user?
2. Top 10 cities that have the highest amount of electricity consumption?
   Which areas has the highest number of smart meters ?
3. One city that has maximum number of electricity connections present and one that consume the highest amount of electricity, are they related ?
4. Compare which energy provider has the maximum connections present and delivers the highest amount of electricity and which has the minimum?

Number of rows(observations) lying between 80,000 to 120,000 represent the connection for zip codes registered under respective network administrators for the specific year with every entry describing at least 10 connections.

In this analysis I only considered electricity dataset of past ten years so that the Jupyter in virtual machine is accessible.

Each file comprises of fourteen distinct features as follow:

1- **Net_manager :** Network administrators region wise (Enexis, Liander, Stedin)

2- **purchase_area :** Area involving energy purchase is designated with an area code

3- **street :** Street name

4- **zipcode_from :** Zipcode of range covered

5- **zipcode_to :** Zipcode of range covered

6- **city :** City Name

7- **num_connections :** Number of connections in the range of zipcodes

8- **delivery_perc :** Electricity or gas consumed%. This number will decrease provided consumer gives back the energy to the grid

9- **perc_of_active_connections :** Active connections in the zipcode range%

10- **type_of_connection :** Main type of connection in the zipcode range. Electricity: No of fuses times amount of Amps.

11- **type_conn_perc :** Availability of main type of connection in the zipcode range%

12- **annual_consume :** Annual consumption. Electricity (kwh).

13- **annual_consume_lowtarif_perc :** Consumption during the low tariff hours%(10 p.m. to 7 a.m. and during weekends)

14- **smartmeter_perc :** Percentage of smartmeters in the zipcode ranges.

Steps:

- Setup Pyspark Environment
- Created Dataframe
- Print Schema and counted the number of rows

```
root
|-- net_manager: string (nullable = true)
|-- purchase_area: string (nullable = true)
|-- street: string (nullable = true)
|-- zipcode_from: string (nullable = true)
|-- zipcode_to: string (nullable = true)
|-- city: string (nullable = true)
|-- num_connections: string (nullable = true)
|-- delivery_perc: string (nullable = true)
|-- perc_of_active_connections: string (nullable = true)
|-- type_conn_perc: string (nullable = true)
|-- type_of_connection: string (nullable = true)
|-- annual_consume: string (nullable = true)
|-- annual_consume_lowtarif_perc: string (nullable = true)
|-- smartmeter_perc: string (nullable = true)
```

This DataFrame has **1710916 rows**.

- Created a column “consumption per connection” by dividing annual consumption with number of active connections, in order to check the amount of electricity consumed by each connection and later calculated the highest consumed amount .

```
+-------------------------+--------------+---------------+
|consumption_per_connection|annual_consume|num_connections|
+-------------------------+--------------+---------------+
|        460.80555555555554|       16589.0|           36.0|
|         583.4615384615385|       22755.0|           39.0|
|        211.89473684210526|        4026.0|           19.0|
|        167.57894736842104|        3184.0|           19.0|
|                    1559.0|       21826.0|           14.0|
```

Maximum amount of consumption by a connection is 9229.666666666666

- Used groupby and orderby function using filters to show top ten cities with highest consumption

```
+----------+--------------------+
|      city|top_ten_consumptions|
+----------+--------------------+
|    MEPPEL|             9999.85|
|      BEST|             9999.78|
|ROOSENDAAL|             9999.66|
|  ROSMALEN|             9999.57|
|  RADEWIJK|             9999.45|
|       OSS|             9999.29|
|    BAEXEM|              9999.0|
| HEEMSKERK|              9999.0|
|   TER AAR|                9999|
|OKKENBROEK|                9999|
+----------+--------------------+
```

- Segregated area into cities and streets and calculated the top 10 areas (cities and streets) which has the highest number of smart meters present.

```
+-------------+--------------------+-------------+
|         city|              street|top_ten_areas|
+-------------+--------------------+-------------+
|    AMSTERDAM|Derde Amstelvlietpad|        99.74|
|    EINDHOVEN|          Torenallee|        99.67|
|    AMSTERDAM|   Bert Haanstrakade|        99.67|
|    AMSTERDAM|Mary van der Slui...|         99.6|
|   AMSTELVEEN|      Maimonideslaan|        99.56|
|'S-GRAVENHAGE|       Waldorpstraat|        99.55|
|       DIEMEN| Carel Willinkgracht|        99.53|
|       ALMERE|          Zeeduinweg|        99.49|
|      UTRECHT| Louis Armstronglaan|        99.47|
|      UTRECHT|       Parkzichtlaan|        99.46|
+-------------+--------------------+-------------+
```

- Used aggregation function to calculate the maximum and minimum number of connections and delivery percentage of electricity in cities and calculated which energy distributor provides the highest and lowest amount of electricity and compared is it related to the number of connections

```
+----------------+--------------+--------------+----------------      -
-+--------------+--------------+
|      net_manager|min_connections|max_connections|avg_no_connections|min_
delivery_by|max_delivery_by|
```

```
+----------------+--------------+--------------+------------------+----
----------+--------------+
|    8716946000005|            10|            88|19.857443609022557|
100.0|          99.21|
|    8716886000004|            10|            99| 23.36839298597823|
0.0|           99.1|
|    8716921000006|            10|            99| 23.24965021919597|
0.0|          99.46|
|    8716924000003|            10|            99|24.711040113596024|
100.0|          99.32|
|    8716874000009|            10|            99|23.079109538090503|
0.0|          99.83|
|    8716925000002|            10|            99|25.629728020240353|
10.71|          99.61|
|    8716892000005|            10|            99|26.913580744743292|
0.0|          99.78|
|           Enexis|             0|            99| 96.20637865535825|
10|             99|
|       Liander NB|          10.0|          99.0|24.175524519496125|
0.0|          99.84|
|Liander N.V. (ZW)|            10|           991| 23.74733308170603|
1.96|           99.8|
|Liander N.V. (NW)|            10|            99|25.415746174082887|
0.0|          99.84|
|     Liander N.V.|            10|            99|22.291289630301556|
0.0|           99.5|
|      Enexis B.V.|             0|            99| 61.17885635973989|
1.09|           99.6|
+----------------+--------------+--------------+------------------+----
----------+--------------+
```

```
+--------------------+--------------------+
|max(max_connections)|min(min_connections)|
+--------------------+--------------------+
|                 991|                   0|
+--------------------+--------------------+
```

Maximum number of electricity connections

```
+-----------------+
|      net_manager|
+-----------------+
|Liander N.V. (ZW)|
+-----------------+
```

Minimum number of electricity connections

```
+-----------+
|net_manager|
+-----------+
|     Enexis|
|Enexis B.V.|
+-----------+
```

Maximum delivery percentage calculated is 99.84

```
+-----------------+
|      net_manager|
+-----------------+
|       Liander NB|
|Liander N.V. (NW)|
```

Minimum delivery percentage calculated is 0.0

```
-----------------+
|      net_manager|
+-----------------+
|    8716886000004|
|    8716921000006|
|    8716874000009|
|    8716892000005|
|       Liander NB|
|Liander N.V. (NW)|
|     Liander N.V.|
+-----------------+
```

CONCLUSIONS

Q1.What the consumption of electricity per user?

Electricity consumption by each user is calculated above and the maximum amount of consumption done by a single user is 9229.666666666666 kwh.

Q2.Top 10 cities that have the highest amount of electricity consumption?

List of top 10 cities that consume the highest amount of electricity is shown above. Meppel consumes the most 9999.85 kwh.

Q3.Which areas has the highest number of smart meters ?

City Amsterdam and street Derde Amstelvlietpad has the highest number of smart meters. List of top 10 areas is shown above.

Q4.One city that has maximum number of electricity connections present and one that consume the highest amount of electricity, are they related ?

Meppel consumes the highest amount of electricity in Netherland but Gerkesklooster has the highest number of electricity connection, therefore its not necessary that if a city has huge number of connection than the consumption is also high.

Q5. Compare which energy provider has the maximum connections present and delivers the highest amount of electricity and which has the minimum?

Maximum number of connections and highest delivery percentage is of Liander N.V but Enexis has the minimum number of connections but its delivery is not the least, therefore delivery percentage of energy distributors is not dependent on the number of connections they have.