LPG Sale Analysis Based on Consumer Demand and Price Forecast

LPG is a low carbon footprint energy source. It constitutes of Propane and Butane. Propane, which comes from natural gas processing and oil refining, is a flammable hydrocarbon gas that is liquefied through pressurization. It is commonly used for heating and cooking, but can be used for a wide range of residential and commercial uses. The percentage of Propane and Butane in LPG differs for the usage area. The usage of LPG differs from country to country. In some countries the big portion of usage is for heating, in those type of countries the biggest portion in demand is for Propane. Also it is used in the production of petrochemical products. In this category China is the biggest importer of LPG in the world followed by India. LPG demand can be categorized into three area:

1. Autogas LPG: Some cars containing the necessary mechanism use LPG as a fuel source. These type of cars are widely common in Turkey and Russia. In this category Turkey is among the biggest LPG consumer in the autogas LPG.
2. Cylindrical: LPG is stored in a small metal container with varying size. It is mostly used for cooking. India is the biggest consumer in this category.
3. Bulk LPG: In this category it is used for heating of buildings. Europe is the biggest consumer in this category.

In this article we will be analyzing LPG usage in each sub category and try to make a connection between each category and public holidays in Turkey. We hope to answer the following questions.

1. Which holidays significantly effect the LPG consumption in each category?
2. What is the relationship between number of Covid-19 cases and LPG usage in each category?
3. Which features affect the LPG consumption and which machine learning model is the best fit to predict the LPG consumption in each category

Now we have a long way to go. Let's begin with exploring our data and try to capture some statistical analysis based on the data on hand. All the analysis and coding is accessible through GitHub.

```
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Cylindrical Sale              588 non-null    int64
 1   Bulk Sale                     588 non-null    int64
 2   Autogas Sale                  588 non-null    int64
 3   Limit                         588 non-null    float64
 4   TUPRAS Price Change  TL/ton   588 non-null    float64
 5   TUPRAS Price (OTV+Gelir)      588 non-null    float64
 6   Year                          588 non-null    int64
 7   Month                         588 non-null    int64
 8   Day                           588 non-null    int64
 9   Weekday                       588 non-null    int64
 10  Special_Day_Type_Holiday      588 non-null    object
 11  Special_Day_Type_Month        588 non-null    object
 12  Special_Day_Type_Raise        588 non-null    object
 13  Special_Day_Type_Discount     588 non-null    object
 14  Covid Cases                   588 non-null    int64
 15  Covid Deaths                  588 non-null    int64
 16  Death/recovering              588 non-null    float64
```

Historical LPG sales data in each category and other features we hope to make connection in the demand quantity of each category

In our dataset we 4 important categorical columns as follows:

1. Special_Day_Type_Holiday: Categorization of national and religious holidays in Turkey.

Categorical Variable Explanation for the national and religious days in Turkey:

Y-1: Previous Day of the new year,

Y: New year ,

Y+1: The day after new year,

T: National sovereign and children's day, Labor and Solidarity day, Commemoration of Atatürk, Youth and Sports Day, Democracy Day, Victory day, Republic Day,

RBA: The day before aid-al-fitr,

RB1: aid-al-fitr day 1,

RB2: aid-al-fitr day 2,

RB3: aid-al-fitr day 3,

KBA: The day before aid-al-adha,

KB1: aid-al-adha day 1,

KB2: aid-al-adha day 2,

KB3: aid-al-adha day 3,

KB4: aid-al-adha day 4,

RBA-2: Three days before the aid-al-fitr,

RBA-1: Two days before the aid-al-fitr,

RB+1: The day after the aid-al-fitr,

RB+2: Two days after the aid-al-fitr,

KBA-2: Three days before the aid-al-adha,

KBA-1: Two days before the aid-al-adha,

KB+1: The day after the aid-al-adha,

KB+2: Two days after the aid-al-adha.

2. Special_Day_Type_Month:

A0: First day of the month

A-1: Last day of the month

A+1: The second day of the month

LPG prices are based on the Sonatrach, Algerian state owned LPG company main supplier of LPG market in Turkey, term price. Sonatrach term price is being updated every month. LPG prices in Turkey change based on a mechanism. The main inputs of the mechanism are TL/USD rate and Sonatrach term price. Days before and after the price change is an important feature in the consumer demand. So we categorized the days before and after the price increase or decrease.

3. Special_Day_Raise:

Z: Raise day

Z+1: The day after the raise day

Z-1: The day before the raise day

Z-2: Two days before the raise day

4. Special_Day_Discount:

I: Price decrease day

I+1: The day after the price decrease day

I-1: The day before the price decrease day

I-2: Two days before the price decrease day

In order to see the sale trends based on the previous day and the previous week we added new columns by shifting the sales and dropped the NA values.

```
data["Autogas-7"] = data["Autogas Sale"].shift(7)
data["Autogas-1"] = data["Autogas Sale"].shift(1)


data["Cylindrical-7"] = data["Cylindrical Sale"].shift(7)
data["Cylindrical-1"] = data["Cylindrical Sale"].shift(1)

data["Bulk-7"] = data["Bulk Sale"].shift(7)
data["Bulk-1"] = data["Bulk Sale"].shift(1)
```

Now lets analyze our data to answer our questions mentioned above.

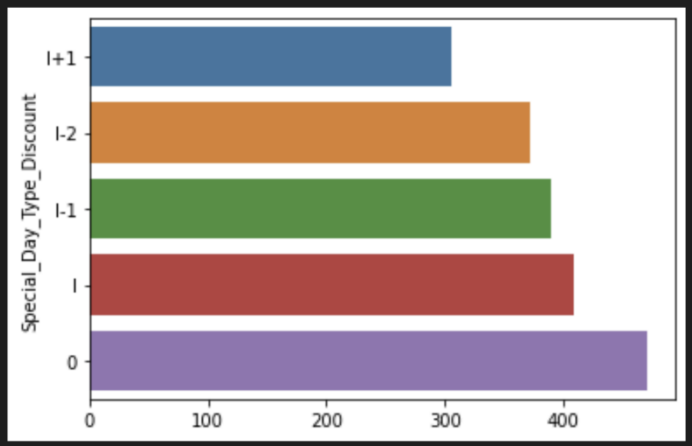
What is the relationship between autogas sales and the discount type, also does the discount or increase effect the sale quantity?

When we look at the mean autogas sales , we see that there is a relationship between autogas sales and discount. Obviously consumers mostly buy autogas in the discount day and the day after.

```
series1 = data.groupby("Special_Day_Type_Discount").mean()['Autogas Sale'].sort_values()
sns.barplot(x=series1.values,y=series1.index)
```

✓ 0.1s

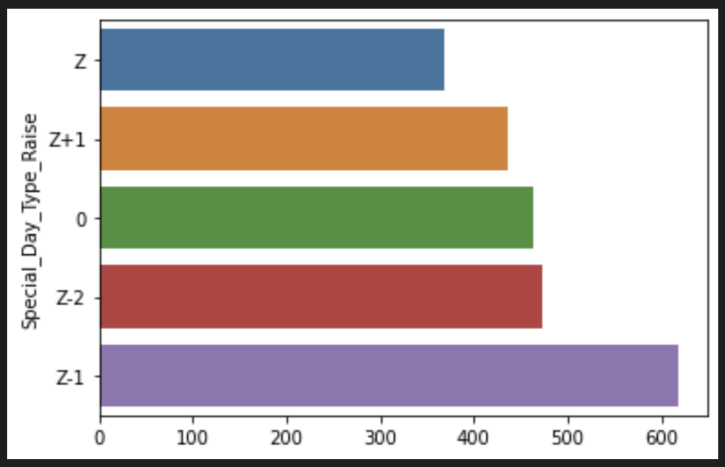
```
<AxesSubplot:ylabel='Special_Day_Type_Discount'>
```



As opposed to the relationship with discount , consumers buy more autogas before the increase.

```
series2 = data.groupby("Special_Day_Type_Raise").mean()['Autogas Sale'].sort_values()
sns.barplot(x=series2.values,y=series2.index);
```

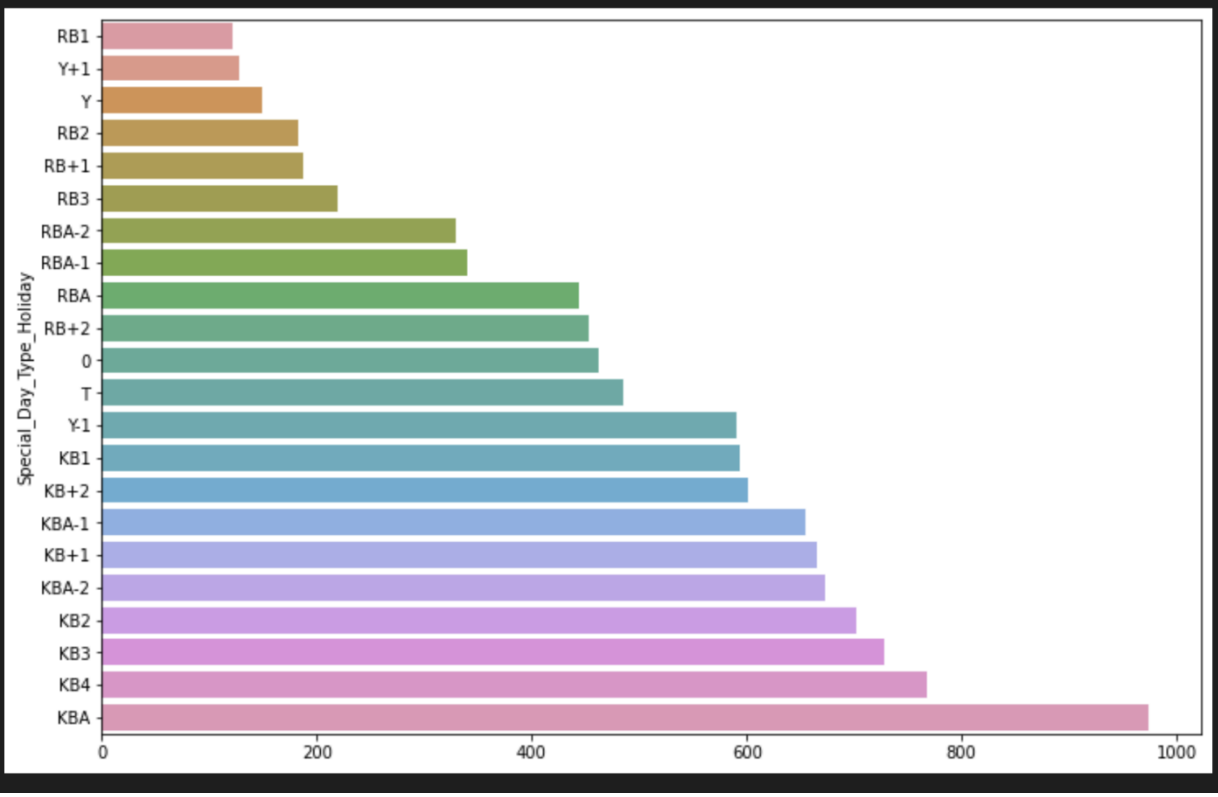
✓ 0.2s



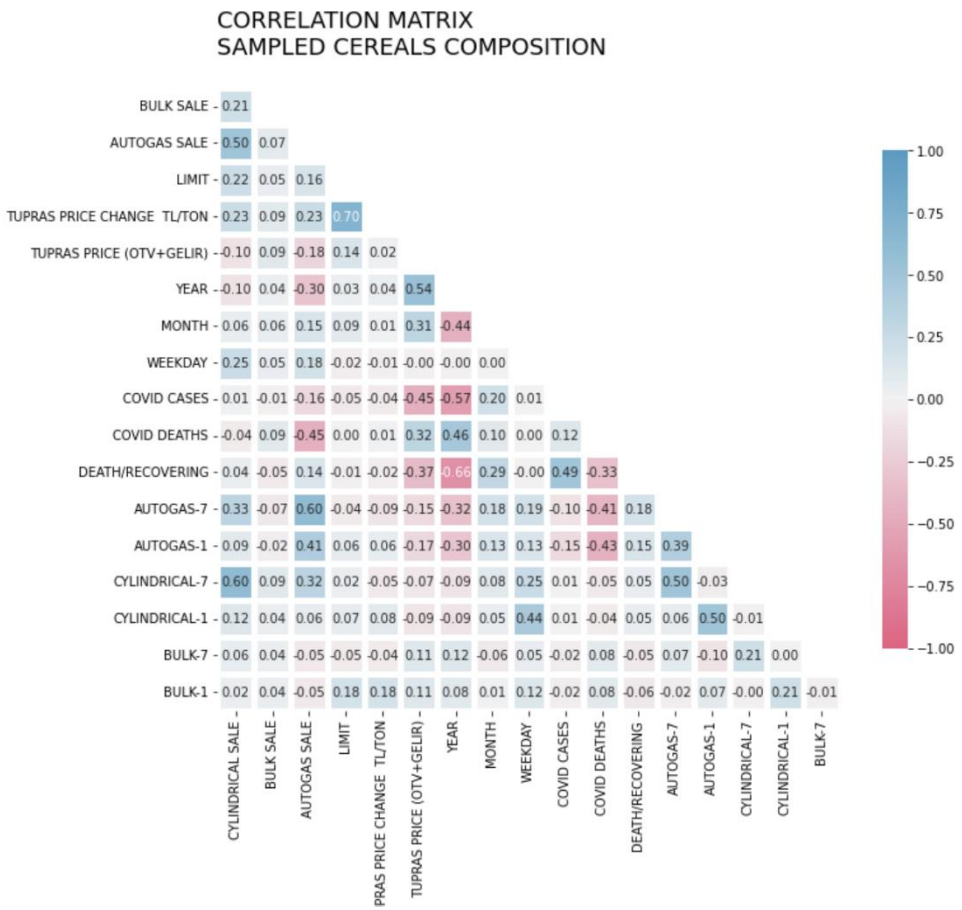
When we analyze the results we observe that consumers have a tendency to buy more autogas before the feast of Eid-al-Adha holiday. Eid-al-Adha holiday is 4 days in Turkey and the day before the Eid-al-Adha holiday is crucial. Demand during the Eid-al-Adhae holiday is really low compared to the normal days. We see that a scenario similar to Eid-al-Adha is also valid for Eid-al-Fitr. While sales are above normal in the first two days before Eid al-Fitr, a dramatic decrease is observed in sales during the Eid-al-Fitr holiday. Although the sales amount is not as much as the sales before Eid-al-Adha and Eid-al-Fitr holiday, an increase in sales is clearly seen before the new year.

```
fig, ax = plt.subplots(figsize=(12, 8))
series3 = data.groupby(["Special_Day_Type_Holiday"]).mean()['Autogas Sale'].sort_values()
sns.barplot(x=series3.values,y=series3.index);
```

✓ 0.3s



We analyzed the weights of special days in autogas demand, now lets take a look at the correlation table and make some comparison.



correlation table of features

As we can predict when we look at the correlation table we see that there is a strong correlation among the sale quantities in each subsequent sale area which are wholesale, cylindrical sales, bulk LPG sale, autogas and total sale. Each subsequent sale amount is positively correlated with the previous day and the previous week sales.

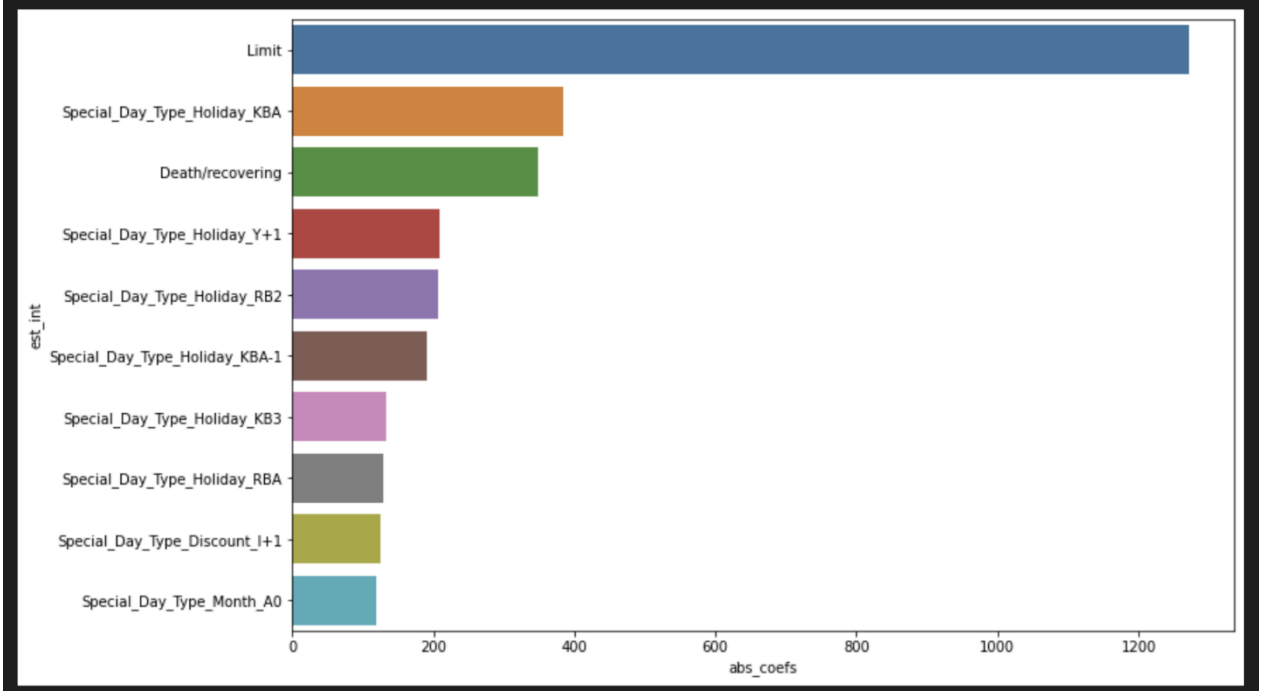
2020 was a special year because of the Covid-19 and related restrictions. There is a significant negative correlation between the number of Covid-19 Cases or Number of Covid-19 Deaths or Death to Recovery Rate and autogas sales but indeed there is a positive correlation for the bulk LPG and cylinder LPG sales. During the increased number of Covid-19 cases people stayed home and come at home and it resulted with the increased number of LPG usage.

What is the importance of each feature in our dataset?

To answer this question we will create a linear model to measure the coefficient weights

```
fig, ax = plt.subplots(figsize=(12, 8))
sns.barplot(x="abs_coefs",y="est_int",data=coef_df.head(10));
```

✓ 0.2s



According to the linear model coefficients, LPG price change plays the biggest role in the prediction of autogas sales. As we discussed in the correlation table we see that previous day of Eid-al-Adha holiday play a very significant role in the prediction of autogas sales.