**Energy Products In KSA**

**Data Insight Group**

**Capstone project report**

**Introduction and problem statement:**

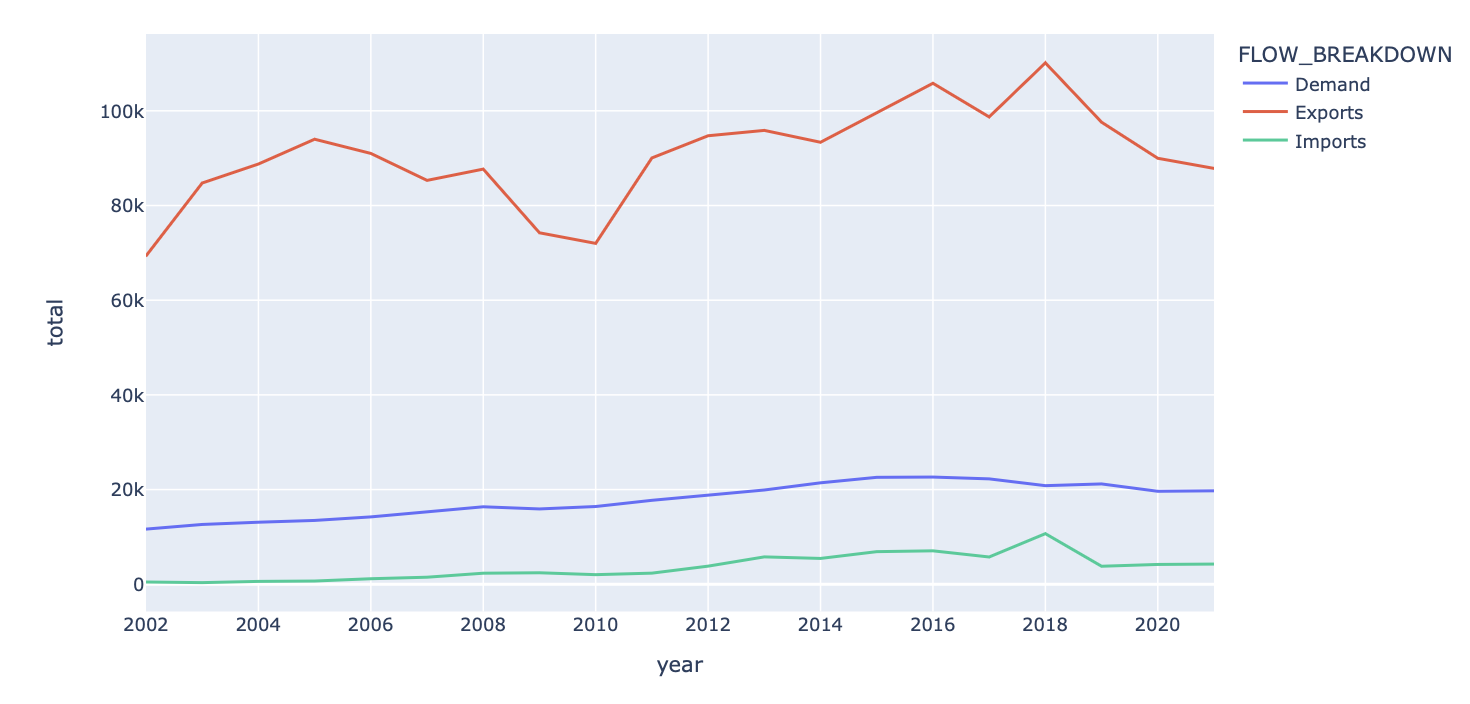
Oil and natural gas are major industries in the energy market and play an influential role in the global economy as the world's primary fuel sources. The processes and systems involved in producing and distributing oil and gas are highly complex, capital-intensive, and require state-of-the-art technology.

Consequently, The Joint Organizations Data Initiative Oil (JODI-Oil) was first launched in 2002, the main goal was not to build a database but to raise the awareness of all oil market players to the need for transparency in oil market data. So, more than 90 countries/economies, Members of the six pioneer organizations (APEC, EUROSTAT, IEA, OLADE, OPEC and UNSD) participate in JODI Oil, representing around 90% of global oil supply and demand. But in our case, we used the date that related to GCC countries.

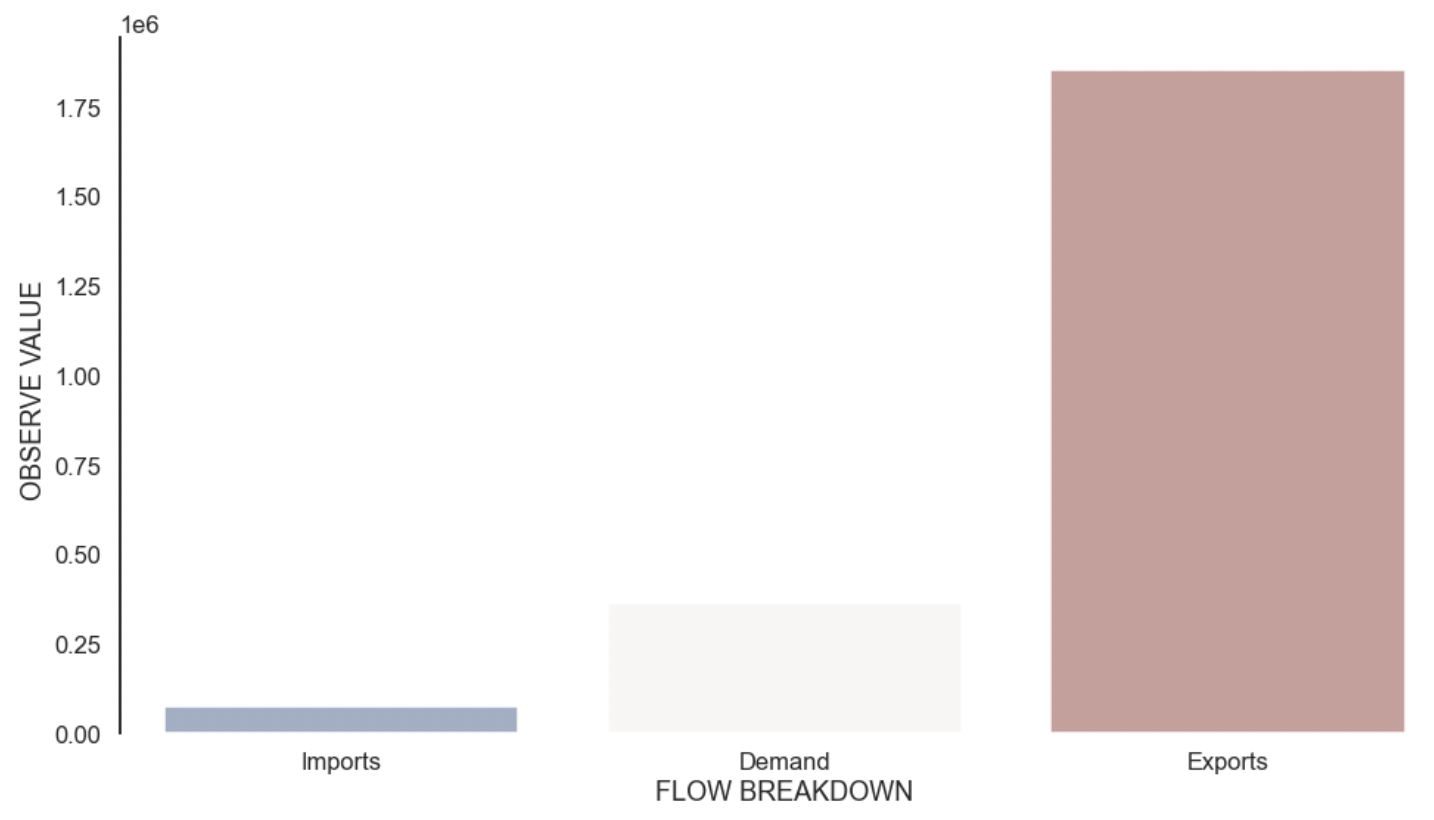
**About the dataset:**

The dataset consists of 955500 rows and seven columns TIME\_PERIOD, AREA, ENERGY\_PRODUCT\_NAME, FLOW\_BREAKDOWN, UNIT\_MEASURE, OBS\_VALUE and ASSESSMENT\_CODE. For the time period it contains data from 2002 till 2021. Moreover, Product categories: crude oil, LPG, gasoline, kerosene, diesel oil, fuel oil and total oil products. Eight flows: production, demand, refinery intake and output, imports, exports, closing stock levels and stock change.

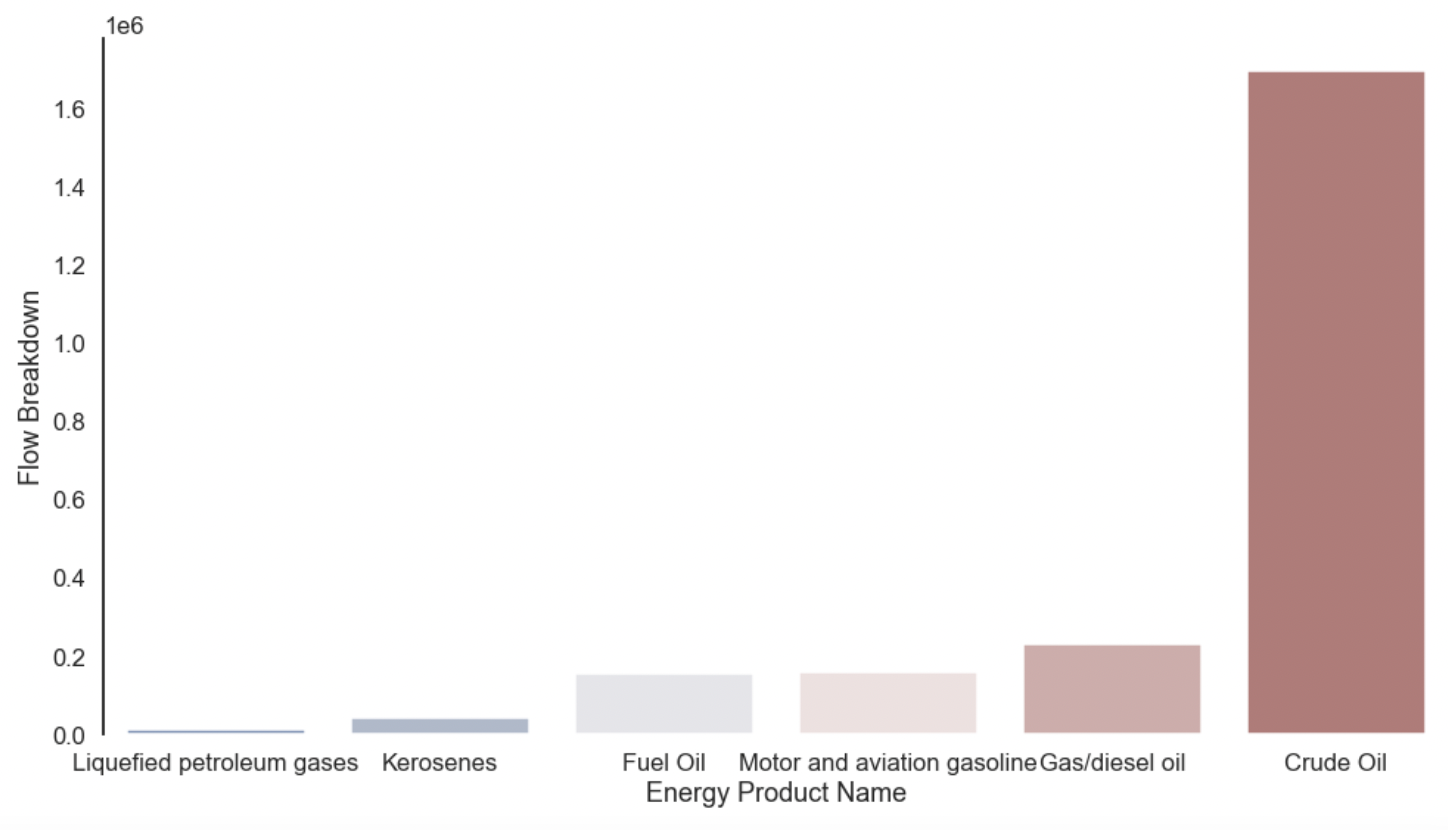
**The explanation of plots:**



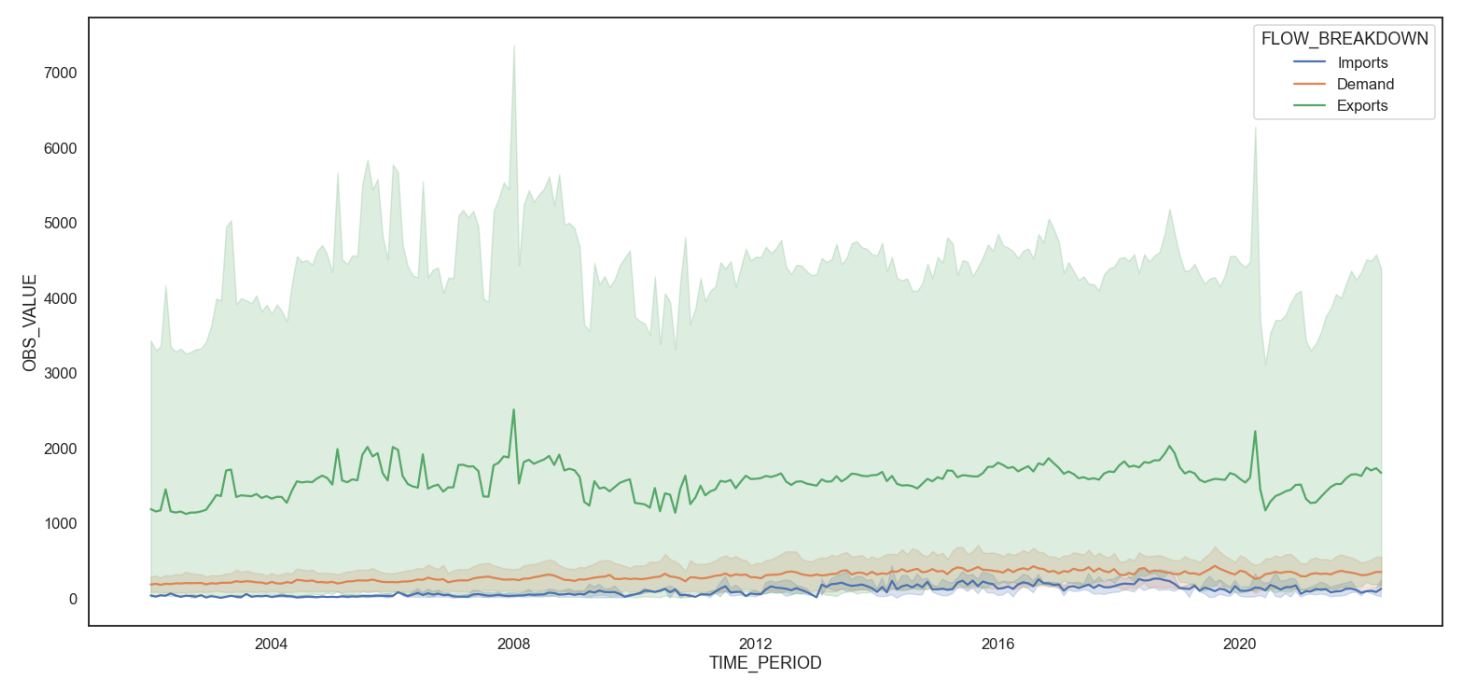
This plot shows the total number of demands, exports, and imports over the years starting from 2002 till 2022, it clearly shows that the exports number starts with more than 60K .



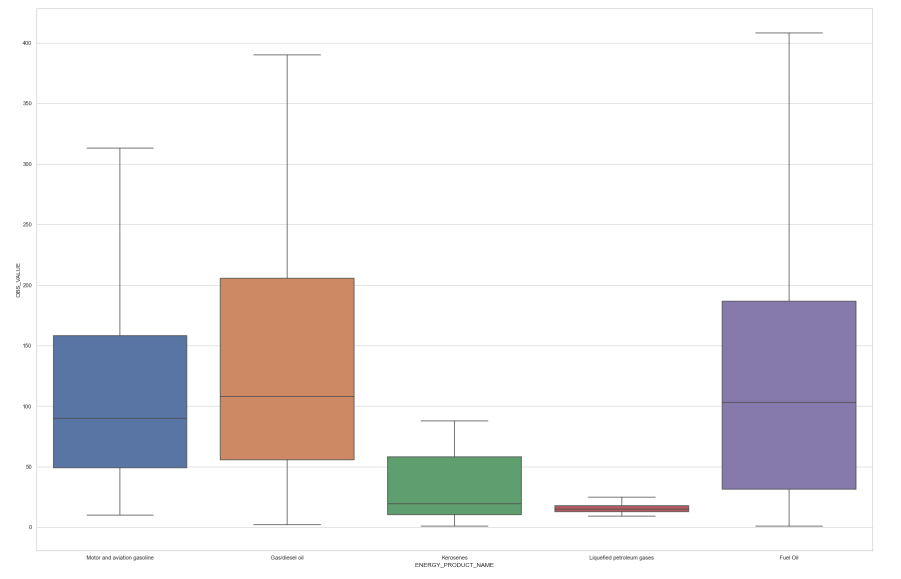
This bar plot represents the counts of demands, exports, and imports that depend on the flow of breakdown. So, imports are close to 700 but for both demands and exports look like the same nearly equal to 1200.



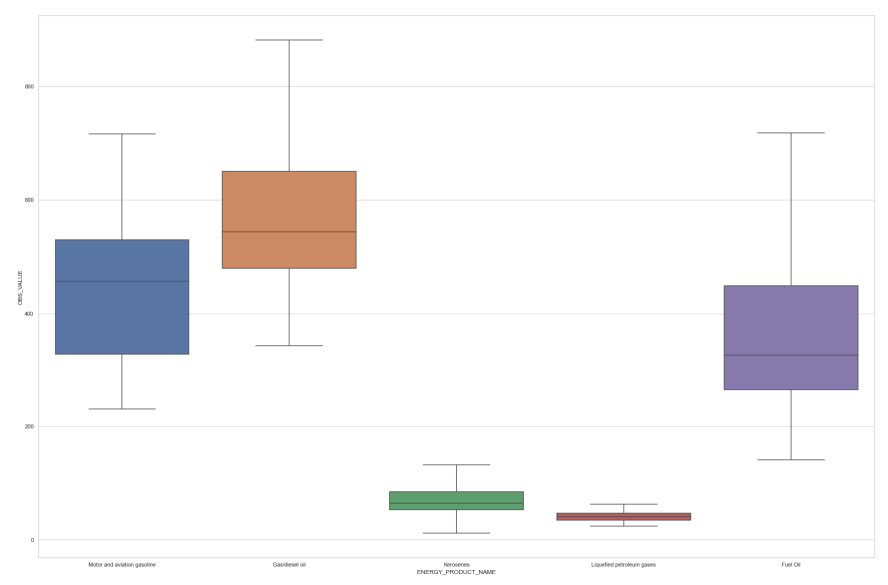
This bar plot shows the total number of the six products: crude oil, liquefied petroleum gasses, motor and aviation gasoline, kerosene, diesel or gas oil, and fuel oil. so that crude oil has the lowest number of products whereas motor and aviation gasoline have the highest number of products compare to others.



For this plot, it shows the OBS value of the imports, exports and demands over the years, so that in 2008, the exports had the most value compared to other years by 7000.

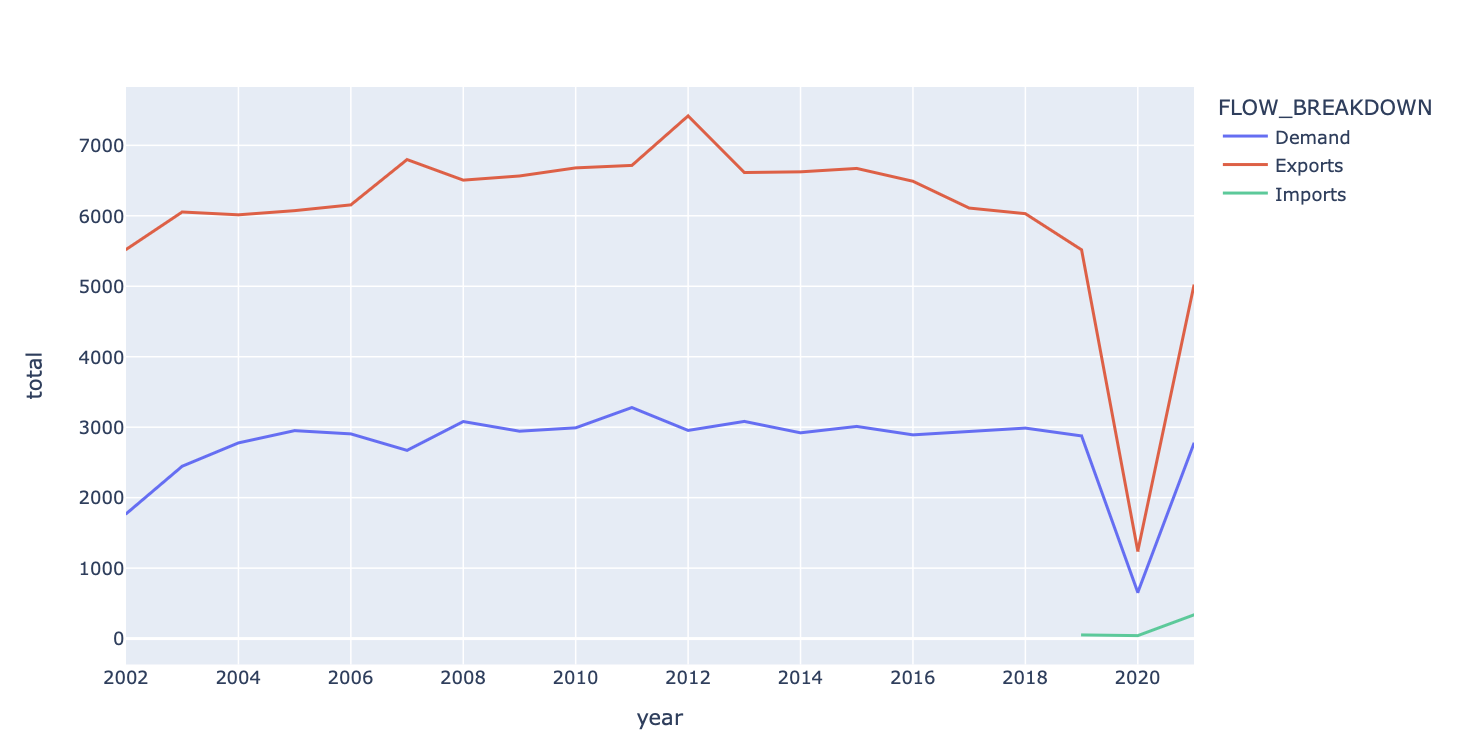


This box plot shows the import in Saudi Arabia for different types of product such as: Fuel Oil, Gas/diesel oil, Kerosenes, Liquefied petroleum gasses and Motor and aviation gasoline. However, it seems that the Fuel Oil has the maximum value with more than 400 thousand, Gas/diesel oil with maximum value near to 390 thousand, Motor and aviation gasoline value is 320 thousand, Kerosenes and Liquefied petroleum gasses these products have less than 100 thousand.

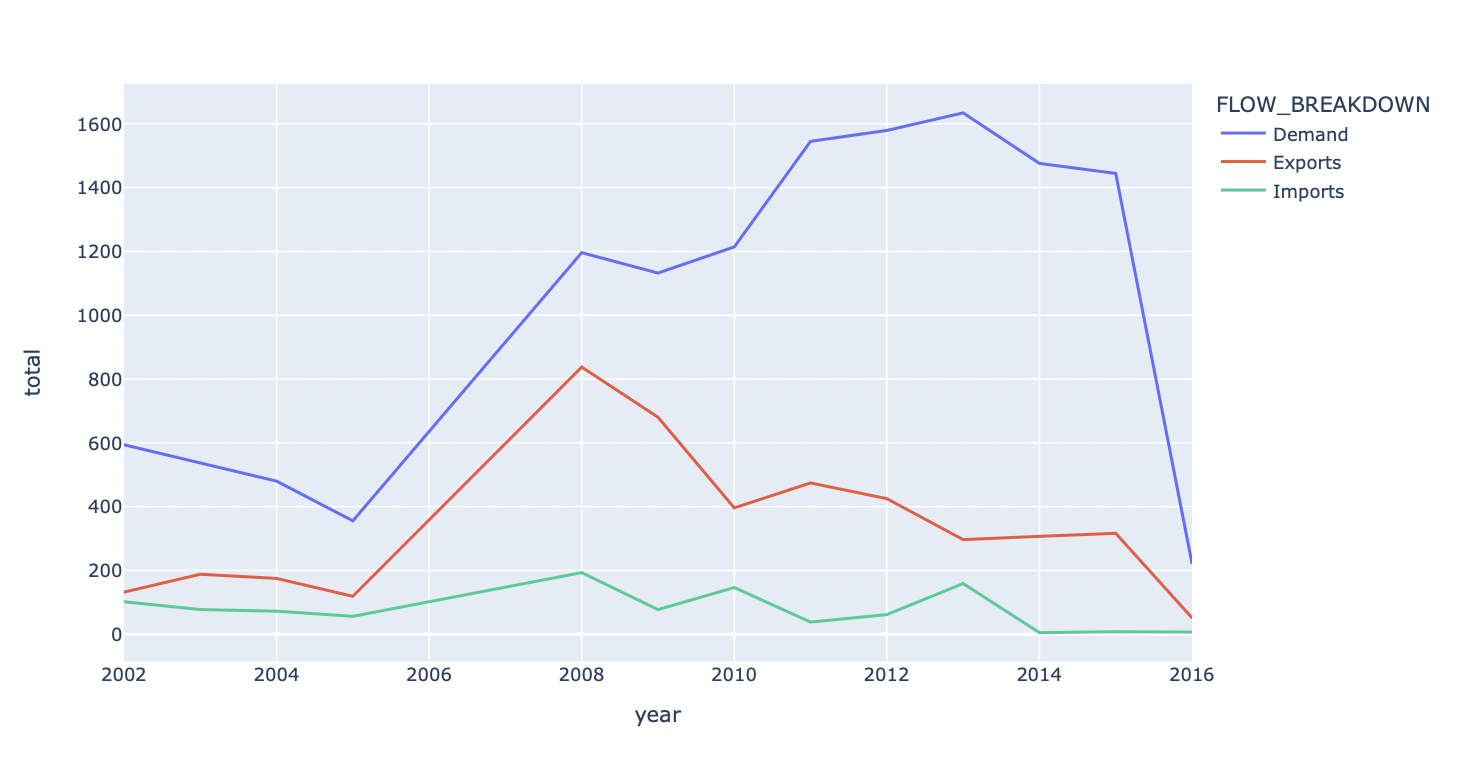


This box plot shows as the previous plot but it investigates the demand products in Saudi Arabia which are: Fuel Oil, Gas/diesel oil, Kerosenes, Liquefied petroleum gasses and Motor and aviation gasoline. However, it seems that the Gas/diesel oil has the highest demand over other products with value near to 900 thousand.

**The next four plots show the total number of demands, exports, and imports over the years of some of the Gulf Cooperation Council countries.**

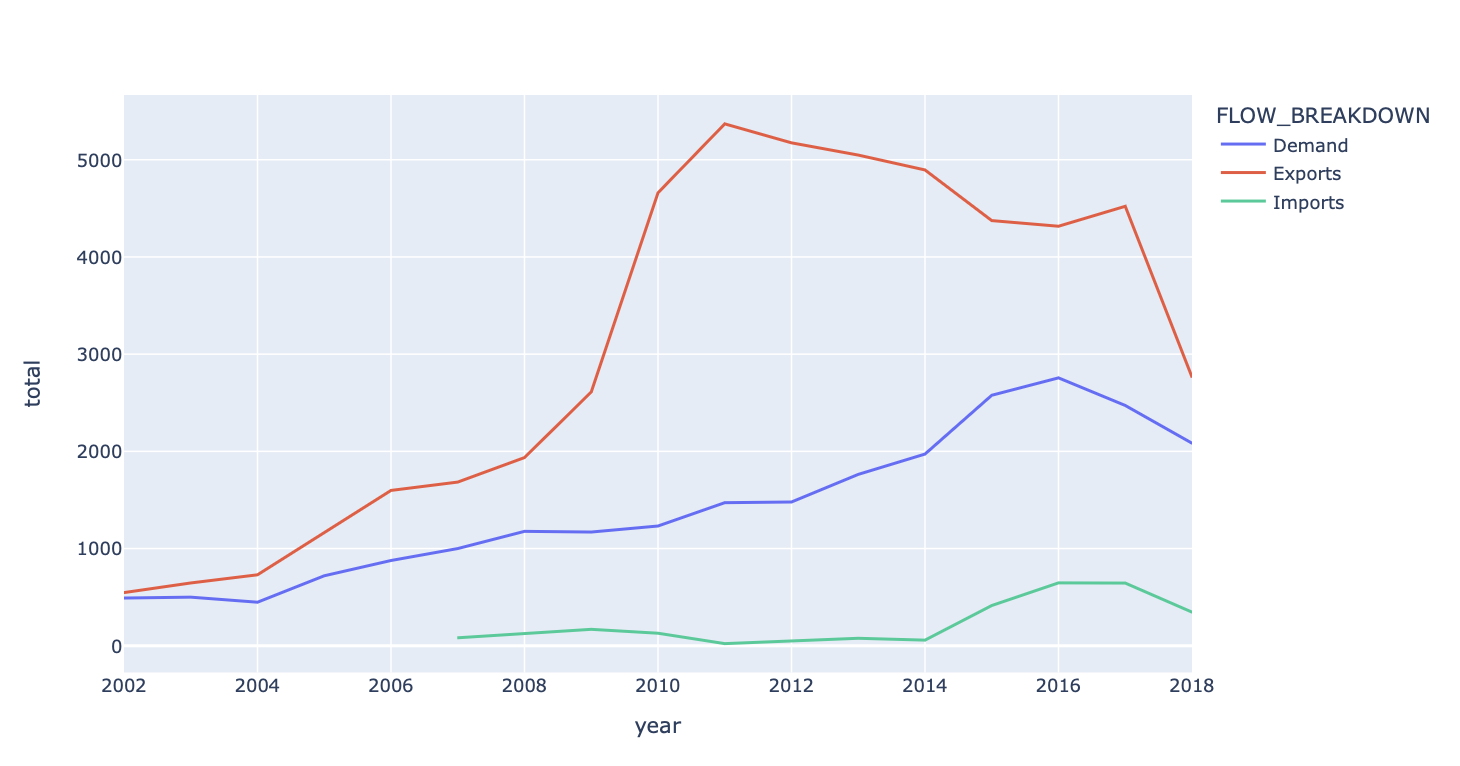


So, this plot represents the Kuwait country's total number of demands, exports, and imports. As shown, the total number of exports will suddenly decrease by nearly 23k in 2020.

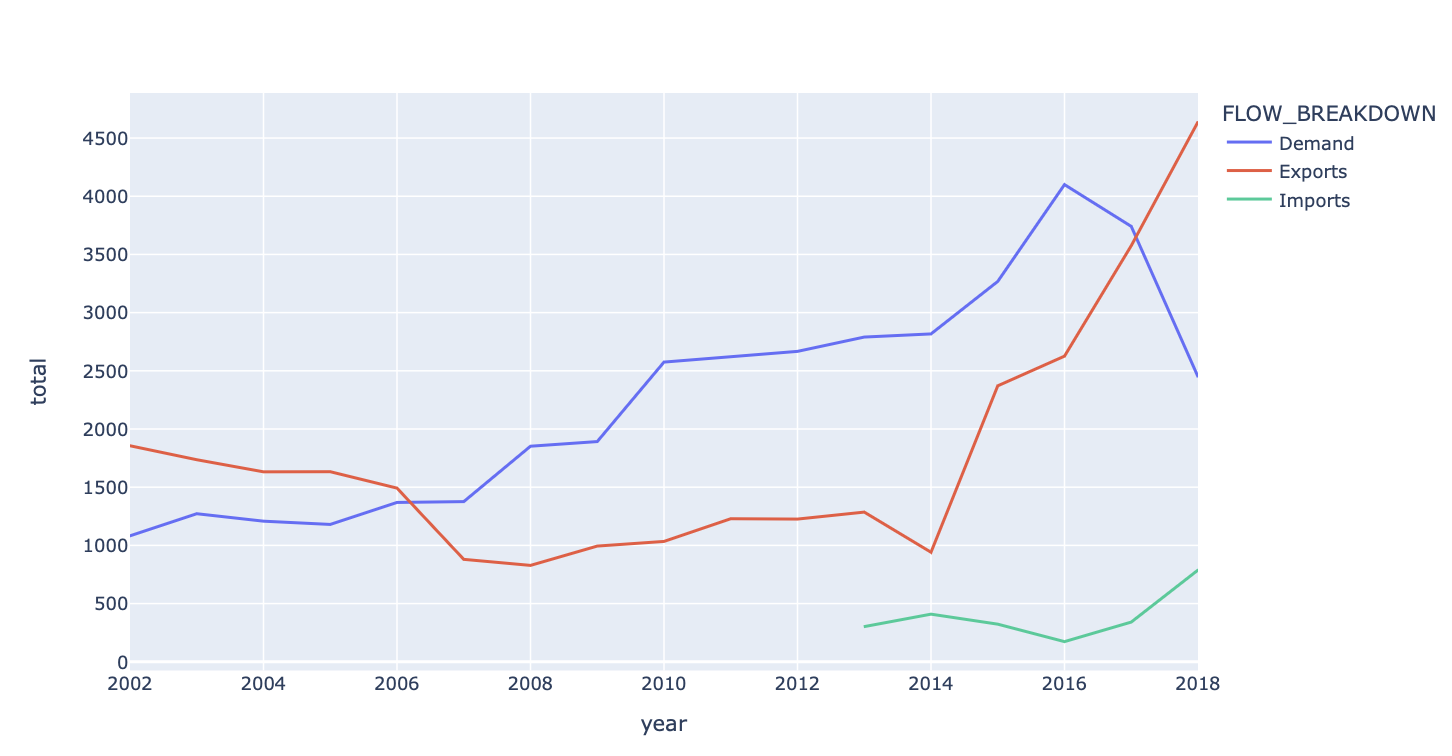


This plot represents the Oman country, which clearly shows the unstable total of exports over the years, in 2005 the total of exports increased from 0 to 8k.

The total number of demands, exports, and imports over the years for Qatar country:



Here is for UAE country:



As shown, the exports total has increased over the years while the demand decreased in 2016.

**Findings and results:**

We use three different models for example: Random Forest, Multiple linear Regression and Polynomial Regression. However, the evaluation for these models is 95% for random forest and 69% and 30% for Polynomial Regression and Multiple linear Regression, respectively. To improve the models and achieve high scores we change the degree parameter for Polynomial Regression from 2 to 4 this modification improves the model to 86%. On the other hand, for Random Forest we use parameter tuning using Grid Search which is a popular technique to enhance the model and achieve 91%.