```
In [1]:  ▶  # Open the image
            from IPython.display import Image
            Image(filename="C:/Users/Rajesh/Downloads/Petroleum.jpg")
```

Out[1]:

To begin this exploratory analysis, first import libraries and define functions for plotting the data using matplotlib. Depending on the data, not all plots will be made.

```
In [2]:  ▶  import pandas as pd    # data processing, CSV file I/O (e.g. pd.read_csv)
            import numpy as np     # linear algebra
            import matplotlib.pyplot as plt    # plotting
            import seaborn as sns  #visualization
            import warnings
            warnings.filterwarnings("ignore")
```

Explore India's Growth in terms of Petroleum Products

```
In [3]:  ▶  data = pd.read_csv("Petroleum Import and Export India Y-o-Y 1998 till 2023.csv")
            data
```

Out[3]:

	IMPORT/EXPORT	PRODUCT	1998-99	1999-2000	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	...	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22 (P)	2022-23 (P)
0	IMPORT	CRUDE OIL	39,808	57,805	74,097	78,706	81,989	90,434	95,861	99,409	...	189238	189435	202850	213932	220433	226498	226955	196461	212382	232732
1	IMPORT	LPG	1,722	1,587	853	659	1,073	1,708	2,334	2,883	...	6567	8313	8959	11097	11380	13235	14809	16476	17043	18309
2	IMPORT	MS	251	0	0	0	0	0	233	486	...	235	372	1012	476	174	670	2146	1351	671	1069
3	IMPORT	Naphtha	2,407	1,917	3,165	3,308	2,784	2,371	2,214	2,331	...	1020	1034	2931	2777	2212	2082	1662	1199	237	897
4	IMPORT	ATF	0	0	0	0	0	0	0	0	...	0	140	286	338	300	259	65	0	0	0
5	IMPORT	SKO	7,065	6,312	1,918	391	698	804	210	1,044	...	0	30	41	0	0	0	0	3	0	0
6	IMPORT	HSD	10,231	5,006	0	31	106	100	814	801	...	77	124	177	1008	1361	555	2796	648	43	328
7	IMPORT	LOBS/ Lube oil	396	407	255	326	340	612	557	1,189	...	2090	2148	2264	2131	2539	2457	2675	2693	3058	2152
8	IMPORT	Fuel Oil	1,696	1,377	1,728	1,977	2,220	1,728	1,585	2,015	...	1331	902	1170	925	1213	1419	4583	6454	8980	8563
9	IMPORT	Bitumen	0	0	0	9	0	6	21	23	...	246	517	879	951	950	877	1630	2055	2581	2787
10	IMPORT	Others	3	1	1,348	308	7	672	860	2,669	...	5130	7722	11735	16585	15333	11794	13423	12369	6403	10437
11	EXPORT	LPG	0	0	0	0	0	0	145	53	...	227	254	195	317	359	417	463	452	513	534
12	EXPORT	MS	0	131	1,202	2,406	2,336	2,979	2,897	2,417	...	15247	16048	16817	15417	14036	12885	12710	11606	13482	13118
13	EXPORT	Naphtha	720	583	2,882	2,535	2,067	2,176	2,926	5,066	...	8322	7008	7116	8727	8951	6963	8897	6509	6861	5714
14	EXPORT	ATF	0	0	160	194	697	1,660	2,480	2,828	...	5745	5520	5686	7271	7183	7389	6906	3544	5186	7264
15	EXPORT	SKO	0	0	0	0	0	0	207	121	...	15	15	10	15	17	19	176	15	14	11
16	EXPORT	HSD	0	0	1,597	2,860	3,178	6,181	7,286	8,504	...	26469	25559	24037	27302	29717	27833	31653	30576	32407	28535
17	EXPORT	LDO	0	0	10	30	0	0	0	0.2	...	30	6	0	151	18	99	0	0	0	1
18	EXPORT	LOBS/ Lube Oil	0	0	0	0	23	17	5	291	...	20	11	17	13	13	8	8	15	10	12
19	EXPORT	Fuel Oil	0	0	508	482	1,120	1,310	1,792	1,815	...	6159	4762	2806	2248	2525	2197	1527	1177	1757	1841
20	EXPORT	Bitumen	0	0	0	0	0	4	47	33	...	95	94	101	38	64	23	25	7	6	9
21	EXPORT	Others	0	32	2,006	1,578	868	293	427	2,333	...	5535	4653	3753	4017	3949	3264	3319	2867	2517	4001

22 rows × 27 columns

To print the information of the data we can use data.info() command.

```
In [4]:  ▶| data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22 entries, 0 to 21
Data columns (total 27 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   IMPORT/EXPORT  22 non-null     object
 1   PRODUCT        22 non-null     object
 2   1998-99        22 non-null     object
 3   1999-2000      22 non-null     object
 4   2000-01        22 non-null     object
 5   2001-02        22 non-null     object
 6   2002-03        22 non-null     object
 7   2003-04        22 non-null     object
 8   2004-05        22 non-null     object
 9   2005-06        22 non-null     object
 10  2006-07        22 non-null     object
 11  2007-08        22 non-null     object
 12  2008-09        22 non-null     object
 13  2009-10        22 non-null     object
 14  2010-11        22 non-null     object
 15  2011-12        22 non-null     int64
 16  2012-13        22 non-null     int64
 17  2013-14        22 non-null     int64
 18  2014-15        22 non-null     int64
 19  2015-16        22 non-null     int64
 20  2016-17        22 non-null     int64
 21  2017-18        22 non-null     int64
 22  2018-19        22 non-null     int64
 23  2019-20        22 non-null     int64
 24  2020-21        22 non-null     int64
 25  2021-22 (P)    22 non-null     int64
 26  2022-23 (P)    22 non-null     int64
dtypes: int64(12), object(15)
memory usage: 4.8+ KB
```

```
In [5]:  ▶| # Changing from Long Format to Wide Format for Better Analysis
         data = pd.melt(data,id_vars=['IMPORT/EXPORT','PRODUCT'],var_name='Fiscal Year',value_name='Volume')
         data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550 entries, 0 to 549
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   IMPORT/EXPORT  550 non-null    object
 1   PRODUCT        550 non-null    object
 2   Fiscal Year    550 non-null    object
 3   Volume         550 non-null    object
dtypes: object(4)
memory usage: 17.3+ KB
```

```
In [6]:  ▶| data
```

Out[6]:

	IMPORT/EXPORT	PRODUCT	Fiscal Year	Volume
0	IMPORT	CRUDE OIL	1998-99	39,808
1	IMPORT	LPG	1998-99	1,722
2	IMPORT	MS	1998-99	251
3	IMPORT	Naphtha	1998-99	2,407
4	IMPORT	ATF	1998-99	0
...	...	...	...	...
545	EXPORT	LDO	2022-23 (P)	1
546	EXPORT	LOBS/ Lube Oil	2022-23 (P)	12
547	EXPORT	Fuel Oil	2022-23 (P)	1841
548	EXPORT	Bitumen	2022-23 (P)	9
549	EXPORT	Others	2022-23 (P)	4001

550 rows × 4 columns

Let's see the mean, count , minimum and maximum values of the data

```
In [7]:  ▶| data.describe()
```

Out[7]:

	IMPORT/EXPORT	PRODUCT	Fiscal Year	Volume
count	550	550	550	550
unique	2	13	25	456
top	IMPORT	LPG	1998-99	0
freq	275	50	22	58

Data Visualization In this section, we will try to understand and compare all columns.

Let's count the columns with different datatypes like Category, Integer, Float.

```
In [8]:  ▶| data.dtypes
```

```
Out[8]: IMPORT/EXPORT    object
        PRODUCT          object
        Fiscal Year      object
        Volume           object
        dtype: object
```

```
In [9]:  ▶| print(f"Number of categorical columns:", len(data.select_dtypes(include='object').columns))
         print(f"Number of integer columns:", len(data.select_dtypes(include='int').columns))
         print(f"Number of float columns:", len(data.select_dtypes(include='float').columns))
```

```
Number of categorical columns: 4
Number of integer columns: 0
Number of float columns: 0
```

```
In [10]:  ▶| # Now, let's have a look at whether this dataset has any null values or not
          data.isnull().sum()
```

```
Out[10]: IMPORT/EXPORT    0
         PRODUCT          0
         Fiscal Year      0
         Volume           0
         dtype: int64
```

```
In [11]: ▶  data.isna().sum()
```

```
Out[11]: IMPORT/EXPORT    0
         PRODUCT          0
         Fiscal Year      0
         Volume           0
         dtype: int64
```

Its a good thing that there is no missing values!

```
In [12]: ▶  # Exploring PRODUCT type
            data.PRODUCT.value_counts()
```

```
Out[12]: LPG              50
         MS               50
         Naphtha          50
         ATF              50
         SKO              50
         HSD              50
         Fuel Oil         50
         Bitumen          50
         Others           50
         CRUDE OIL        25
         LOBS/ Lube oil   25
         LDO              25
         LOBS/ Lube Oil   25
         Name: PRODUCT, dtype: int64
```

```
In [13]: ▶  # Remove the commas in the volume columns
            data['Volume'] = data['Volume'].astype(str).str.replace(',','',regex=True)
```

The Volume is shown as a string. We need to change it to float values. Also, the strings are comma separated. So we will have to first replace the commas.

```
In [14]: ▶  # Change the Volume column from string to Float
            data['Volume'] = data['Volume'].astype(float)
```

```
In [15]: ▶  data_crude = data[data['PRODUCT']=='CRUDE OIL']
```

```
In [16]: ▶  # Bar Plot for Crude Oil
            plt.figure(figsize=(8,4))
            sns.barplot(data_crude,x='Fiscal Year',y='Volume')
            plt.xticks(rotation=90);
```
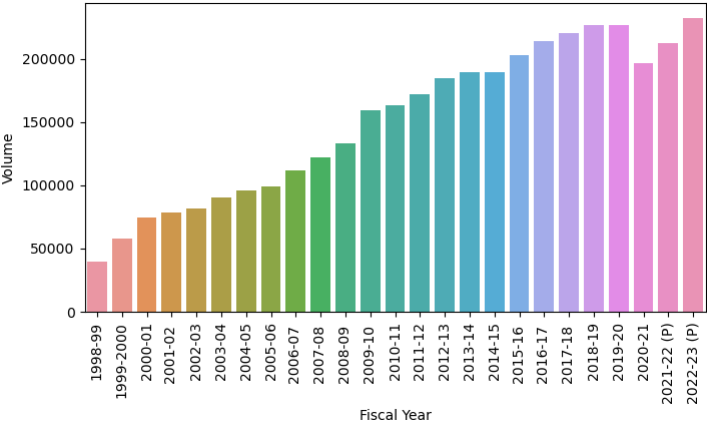

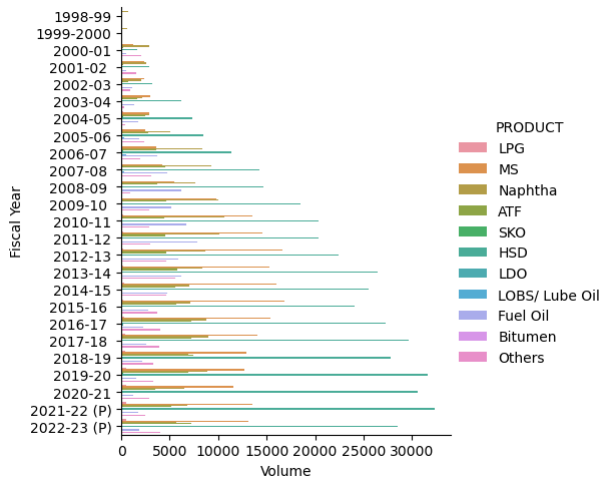Analysis of Exported Petroleum Products from India

```
In [17]: ▶  # Filtering Exported Petroleum Products
            data_products_export = data[(data['PRODUCT']!='CRUDE OIL') & (data['IMPORT/EXPORT']=='EXPORT')]
```

```
In [18]: ▶  plt.figure(figsize=(12,6))
            sns.lineplot(data_products_export,x='Fiscal Year',y='Volume',hue='PRODUCT',size='PRODUCT')
            plt.xticks(rotation=90);
```
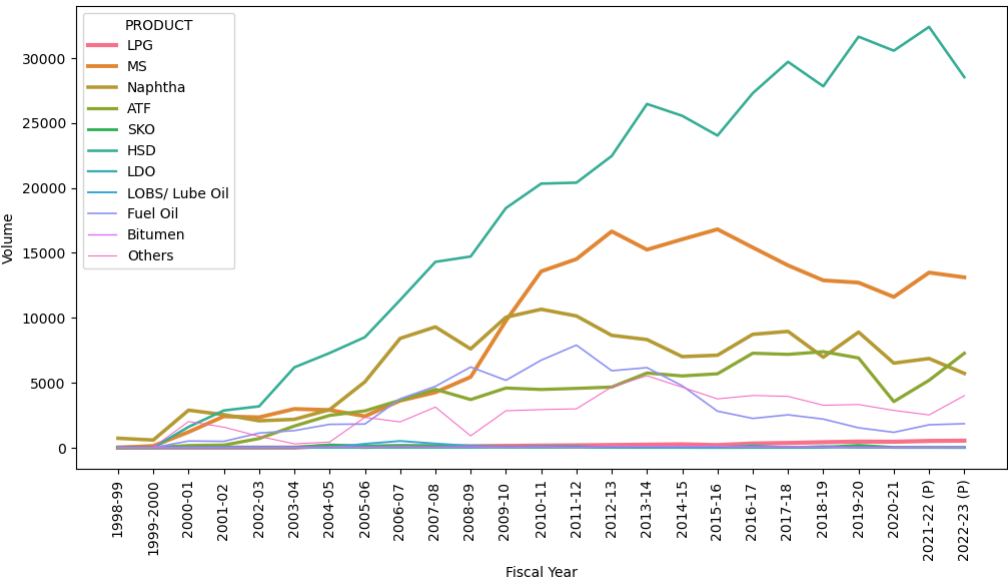
```
In [19]:  sns.catplot(data_products_export,x='Volume',y='Fiscal Year',hue='PRODUCT',kind='bar')
```
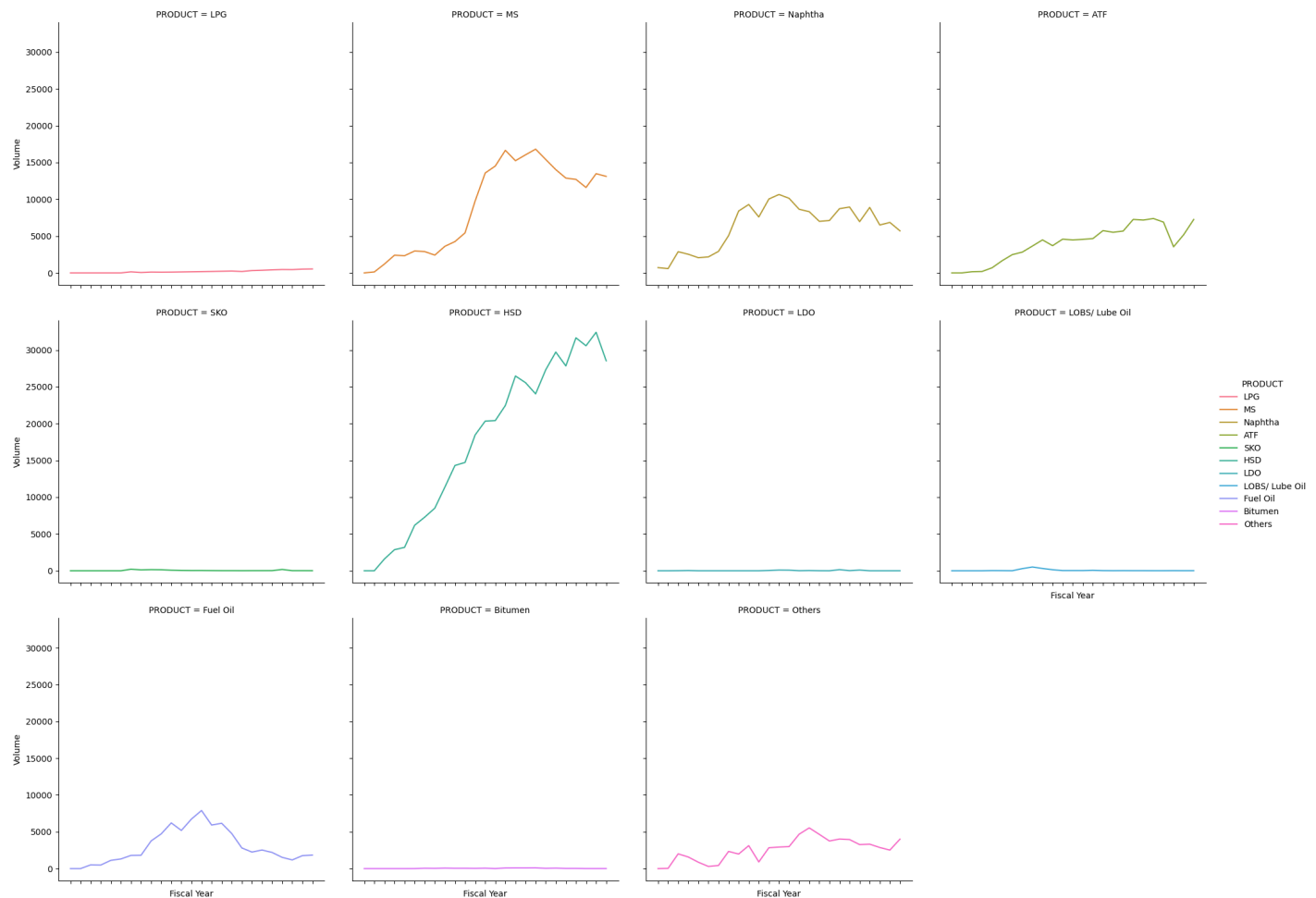
Out[19]: <seaborn.axisgrid.FacetGrid at 0x1d0308304c0>


```
In [20]:  plt.figure(figsize=(12,6))
          fig = sns.relplot(data_products_export,x='Fiscal Year',y='Volume',col='PRODUCT', kind='line',col_wrap = 4,hue='PRODUCT',palette="husl")
          fig.set_xticklabels()
```

Out[20]: <seaborn.axisgrid.FacetGrid at 0x1d0316618d0>

          <Figure size 1200x600 with 0 Axes>

Its evident from the above analysis export of finished or refined petroleum products from India has increased manifolds specially mainstream products like MS(Motor Spirit or Petrol) and HSD (High Speed Diesel).

HSD export growth is exponential

```
Analysis of Imported Petroleum Products
```

```
In [21]:  # Filtering Imported Petroleum Products
          data_products_import = data[(data['PRODUCT']!='CRUDE OIL') & (data['IMPORT/EXPORT']=='IMPORT')]
```

```
In [22]:  ▶  plt.figure(figsize=(12,6))
             sns.lineplot(data_products_import,x='Fiscal Year',y='Volume',hue='PRODUCT')
             plt.xticks(rotation=90);
```



```
In [23]:  ▶  plt.figure(figsize=(12,6))
             fig = sns.relplot(data_products_import,x='Fiscal Year',y='Volume',col='PRODUCT', kind='line',col_wrap = 4,hue='PRODUCT',palette="husl")
             fig.set_xticklabels()
```

Out[23]: <seaborn.axisgrid.FacetGrid at 0x1d030c73dc0>

          <Figure size 1200x600 with 0 Axes>
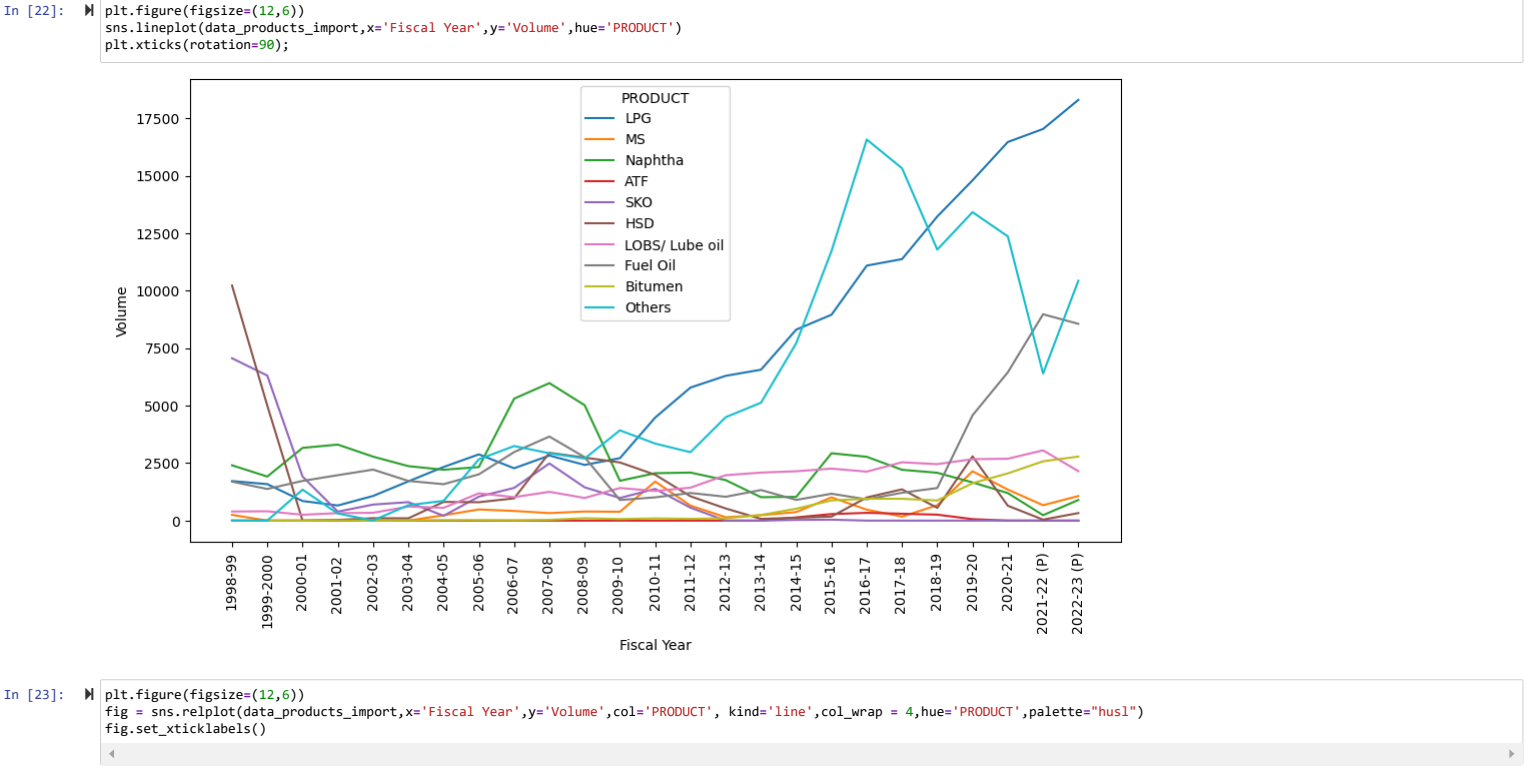
Its clear that mainstream products import has reduced over the years. However, there is steep rise in the import of LPG.

The above exploratory data analysis clearly indicates that India's Growth Story with Petroleum Products.
From importing mainstream products, India has started to export finished products which defines increase in refining capacity.

Rise in import of LPG shows increase in consumption of cleaner fuels.

India has to dedicate its focus on in-house GAS production to sustain the demand.