In [2]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:
```
Export = pd.read_csv('2018-2010_export.csv')
Import = pd.read_csv('2018-2010_import.csv')
```

In [4]:
```
Export.head()
```

Out[4]:

	HSCode	Commodity	value	country	year
0	2	MEAT AND EDIBLE MEAT OFFAL.	0.18	AFGHANISTAN TIS	2018
1	3	FISH AND CRUSTACEANS, MOLLUSCS AND OTHER AQUAT...	0.00	AFGHANISTAN TIS	2018
2	4	DAIRY PRODUCE; BIRDS' EGGS; NATURAL HONEY; EDI...	12.48	AFGHANISTAN TIS	2018
3	6	LIVE TREES AND OTHER PLANTS; BULBS; ROOTS AND ...	0.00	AFGHANISTAN TIS	2018
4	7	EDIBLE VEGETABLES AND CERTAIN ROOTS AND TUBERS.	1.89	AFGHANISTAN TIS	2018

In [5]:
```
(Export.isnull().sum()/ Export.isnull().count() *100).sort_values(ascending=False)
```

Out[5]:
```
value        10.244995
year          0.000000
country       0.000000
Commodity     0.000000
HSCode        0.000000
dtype: float64
```

In [6]:
```
Export['value'].mean()
```

Out[6]:
```
21.567829166156546
```

In [7]:
```
Export['value'].fillna(21.5, inplace=True)
```

In [8]:
```
Export.drop_duplicates(keep="first",inplace=True)
Export['country']= Export['country'].apply(lambda x : np.NaN if x == "UNSPECIFIED" else x)
Export = Export[Export.value!=0]
Export.dropna(inplace=True)
```

In [9]:
```
Import.head()
```

Out[9]:

	HSCode	Commodity	value	country	year
0	5	PRODUCTS OF ANIMAL ORIGIN, NOT ELSEWHERE SPECI...	0.00	AFGHANISTAN TIS	2018
1	7	EDIBLE VEGETABLES AND CERTAIN ROOTS AND TUBERS.	12.38	AFGHANISTAN TIS	2018
2	8	EDIBLE FRUIT AND NUTS; PEEL OR CITRUS FRUIT OR...	268.60	AFGHANISTAN TIS	2018
3	9	COFFEE, TEA, MATE AND SPICES.	35.48	AFGHANISTAN TIS	2018
4	11	PRODUCTS OF THE MILLING INDUSTRY; MALT; STARCH...	NaN	AFGHANISTAN TIS	2018

In [10]:
```
(Import.isnull().sum()/ Import.isnull().count() *100).sort_values(ascending=False)
```

Out[10]:
```
value        15.222532
year          0.000000
country       0.000000
Commodity     0.000000
HSCode        0.000000
dtype: float64
```

In [11]:
```
mean_value = Import.value.mean()
Import.value.fillna(mean_value, inplace = True )
Import.drop_duplicates(keep="first",inplace=True)
Import['country']= Import['country'].apply(lambda x : np.NaN if x == "UNSPECIFIED" else x)
Import = Import[Import.value!=0]
Import.dropna(inplace=True)
```

In [12]:
```
Import.isnull().sum()
```

Out[12]:
```
HSCode       0
Commodity    0
value        0
country      0
year         0
dtype: int64
```

In [13]:
```
Export.loc[Export.value == Export.value.max()]
```

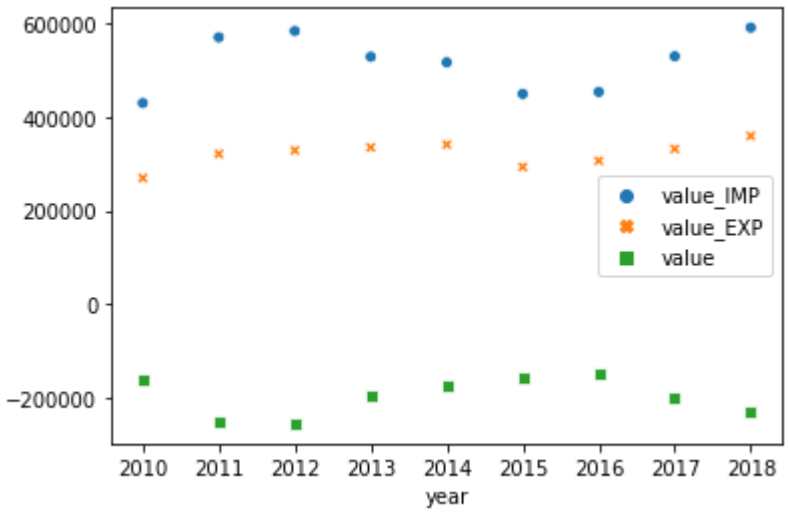
Out[13]:

	HSCode	Commodity	value	country	year
135854	71	NATURAL OR CULTURED PEARLS,PRECIOUS OR SEMIPRE...	19805.17	U ARAB EMTS	2010

Below graph depicts the difference between Export and Import process in India Trade and the difference between them is known as deficit.

In [14]:
```
total_imports_per_year=Import.groupby('year').agg({'value':'sum'})
total_exports_per_year=Export.groupby('year').agg({'value':'sum'})
deficit = Export.groupby('year').agg({'value':'sum'}) - Import.groupby('year')
.agg({'value':'sum'})

t = total_imports_per_year.join(total_exports_per_year, lsuffix='_IMP', rsuffi
x='_EXP')
r = t.join(deficit)
sns.scatterplot(data=r)
```

Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x1a9f3f3d948>



Thing to Notice :-

1)India's biggest deficit was from 2011 to 2012 as imports were more than exports, with a amount in excess of $200,000 million .
2)Then the deficit reappeared in 2014.
3)2016 was the lowest year of deficit

LARGEST IMPORTERS & EXPORTERS OF INDIA 2010-2018

1) I have recogised a group of countries to which India is exporting.

2) Find the sum of export values for each country.

3) I have created a DataFrame containing the country and the sum of the values.

4) Then I identified the top 5 countries.

5) Then I drew the relationship using a bar plot.

6) Then I repeat this steps for importers.

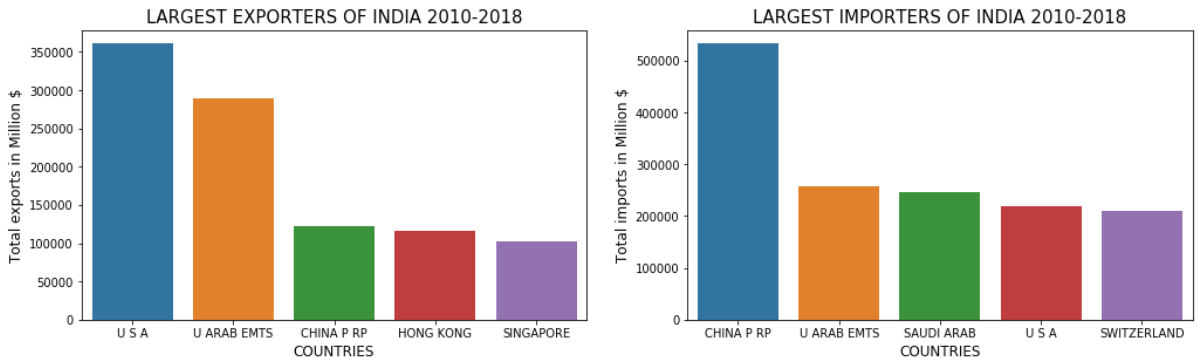
In [15]:
```
Export.describe()
```

Out[15]:

	HSCode	value	year
count	120510.000000	120510.000000	120510.000000
mean	50.757049	23.977592	2014.045830
std	27.744305	225.620344	2.577779
min	1.000000	0.010000	2010.000000
25%	28.000000	0.110000	2012.000000
50%	52.000000	1.110000	2014.000000
75%	73.000000	14.310000	2016.000000
max	99.000000	19805.170000	2018.000000

In [16]:
```
country_export_list=list(Export.country.unique())
country_import_list=list(Import.country.unique())
```

In [17]:
```
country_export_group=Export.groupby('country')
country_import_group=Import.groupby('country')

ls_export=[]
ls_import = []
for country_name in country_export_list:
    ls_export.append([country_name, country_export_group.get_group(str(country
_name)).value.sum() ])
for country_name in country_import_list:
    ls_import.append([country_name, country_import_group.get_group(str(country
_name)).value.sum() ])
total_exports = pd.DataFrame(ls_export, columns = ['country', 'total_exports'
])
total_imports = pd.DataFrame(ls_import, columns = ['country', 'total_imports'
])

largest_exporters_dataframe=total_exports.nlargest(5,['total_exports'])
largest_importers_dataframe=total_imports.nlargest(5,['total_imports'])

fig = plt.figure(figsize=(17,10))
ax = fig.add_subplot(221)
sns.barplot(largest_exporters_dataframe['country'],largest_exporters_dataframe
['total_exports'])
plt.xlabel('COUNTRIES',size=12)
plt.ylabel('Total exports in Million $',size=12)
plt.title('LARGEST EXPORTERS OF INDIA 2010-2018',SIZE=15)
ax = fig.add_subplot(222)
sns.barplot(largest_importers_dataframe['country'],largest_importers_dataframe
['total_imports'])

plt.xlabel('COUNTRIES',size=12)
plt.ylabel('Total imports in Million $',size=12)
plt.title('LARGEST IMPORTERS OF INDIA 2010-2018',SIZE=15)

plt.show()
```



Things to Notice :-

1)The largest exporter is the United States of America, followed by the United Arab Emirates.

2)China has huge market in India followed by UAE,Saudi Arabia and USA(The largest country that India imports to)

Trade Defecit/Surplus:-

China - very Huge Trade Deficit (imports 500m >>exports360m)

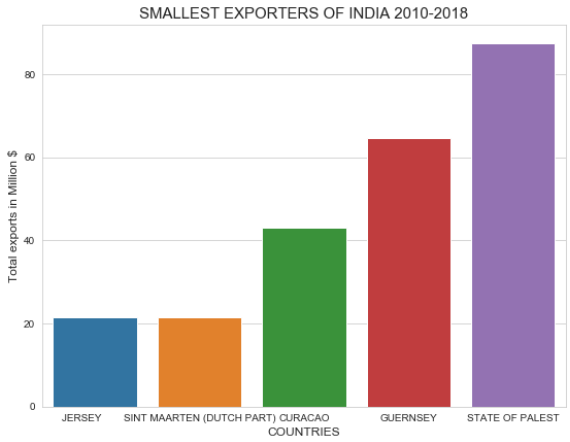
UAE - little Trade Surplus (imports 270m <<exports295m)

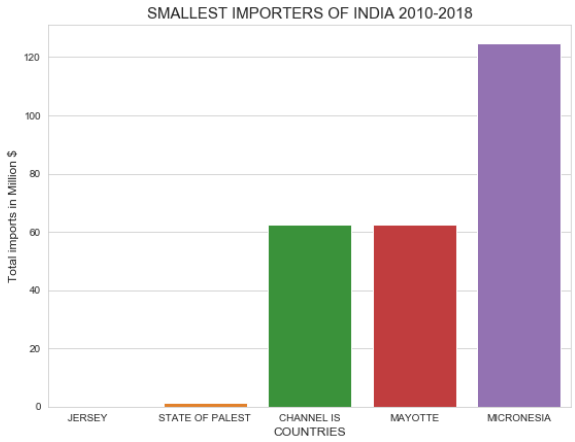
USA - little Trade Surplus (imports 220m <<exports355m)

Smallest Exports of India

In [18]:
```
smallest_exporters_df=total_exports.nsmallest(5,['total_exports'])
smallest_importers_df=total_imports.nsmallest(5,['total_imports'])
```

In [32]:
```
fig = plt.figure(figsize=(20,15))
ax = fig.add_subplot(221)
sns.barplot(smallest_exporters_df['country'],smallest_exporters_df['total_expo
rts'])
plt.xlabel('COUNTRIES',size=12)
plt.ylabel('Total exports in Million $',size=12)
plt.title('SMALLEST EXPORTERS OF INDIA 2010-2018',SIZE=15)
ax = fig.add_subplot(222)
sns.barplot(smallest_importers_df['country'],smallest_importers_df['total_impo
rts'])

plt.xlabel('COUNTRIES',size=12)
plt.ylabel('Total imports in Million $',size=12)
plt.title('SMALLEST IMPORTERS OF INDIA 2010-2018',SIZE=15)

plt.show()
```





Note that:

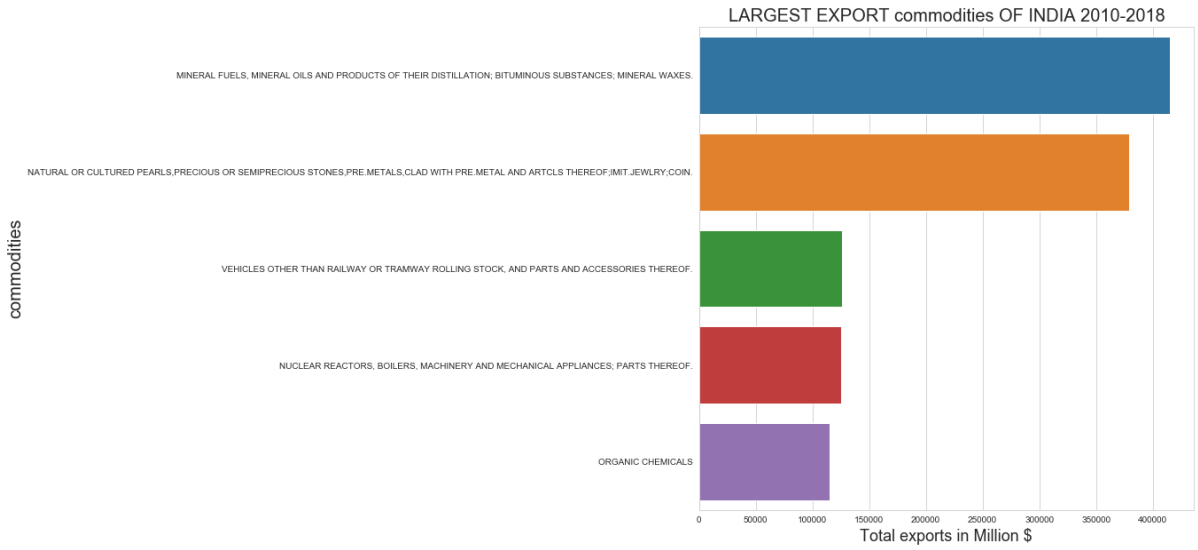
1)The lowest country that India exports from is the Jersey then Sint Maarten.

2)The lowest country that India imports to is the Jersey then Palestine.

Export commodity of India

In [20]:
```
commodity_export_list=list(Export.Commodity.unique())
commodity_export_group=Export.groupby('Commodity')
ls=[]
for commodity_name in commodity_export_list:
    ls.append([commodity_name, commodity_export_group.get_group(str(commodity_
name)).value.sum() ])

total_exports = pd.DataFrame(ls, columns = ['Commodity', 'total_exports'])


largest_commodities_dataframe=total_exports.nlargest(5,['total_exports'])

plt.figure(figsize=(10,10))
sns.set_style('whitegrid')
largest_commodities_bar=sns.barplot(y=largest_commodities_dataframe['Commodit
y'],x=largest_commodities_dataframe['total_exports'])
plt.ylabel('commodities',size=20)
plt.xlabel('Total exports in Million $',size=18)
plt.title('LARGEST EXPORT commodities OF INDIA 2010-2018',SIZE=20)
```

Out[20]: Text(0.5, 1.0, 'LARGEST EXPORT commodities OF INDIA 2010-2018')



Things to notice:-

1) The largest Commodity that India exports is "MINERAL FUELS&OILS".

2) Second largest the "PEARLS".

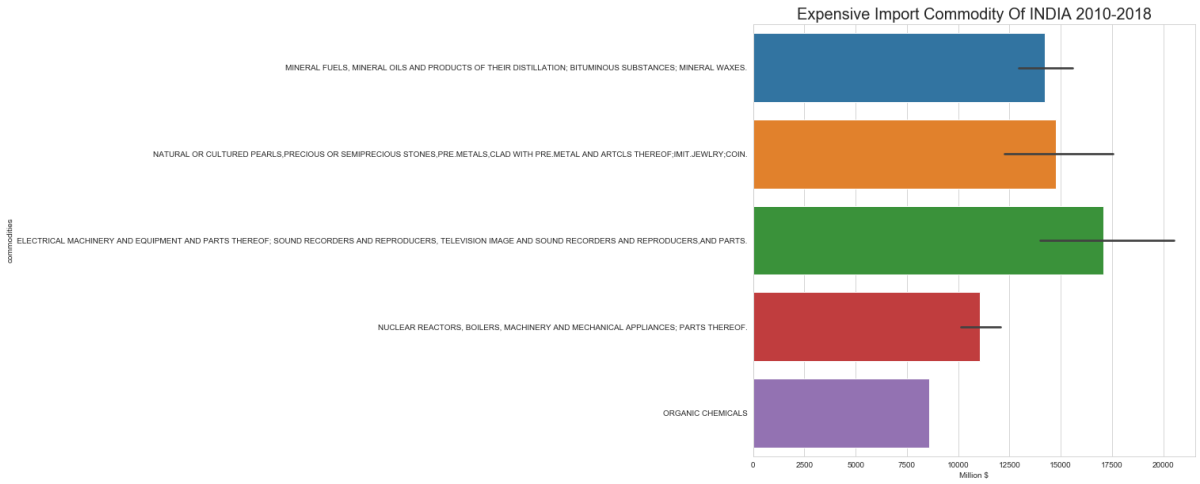
Expensive Import Commodity of India

In [22]:
```
expensive_import = Import[Import.value>3000]
expensive_import5=expensive_import.nlargest(20,['value'])
expensive_import.shape
```

Out[22]: (218, 5)

In [33]:
```
expensive_import = Import[Import.value>1000]
expensive_import5=expensive_import.nlargest(100,['value'])
plt.figure(figsize=(10,10))
sns.set_style('whitegrid')
expensive_import_commodities_bar=sns.barplot(y=expensive_import5['Commodity'],
x=expensive_import5['value'])
plt.ylabel('commodities')
plt.xlabel('Million $')
plt.title('Expensive Import Commodity Of INDIA 2010-2018',SIZE=20)
```

Out[33]: Text(0.5, 1.0, 'Expensive Import Commodity Of INDIA 2010-2018')



Expensive export commodity of India

In [34]:
```
expensive_export = Export[Export.value>1000]
expensive_export5=expensive_export.nlargest(50,['value'])
plt.figure(figsize=(10,10))
expensive_export_commodities_bar=sns.barplot(y=expensive_export5['Commodity'],
x=expensive_export5['value'])
plt.ylabel('commodities')
plt.xlabel('Million $')
plt.title('Expensive Export Commodity Of INDIA 2010-2018',SIZE=20)
```

Out[34]: Text(0.5, 1.0, 'Expensive Export Commodity Of INDIA 2010-2018')