Data Analysis Portfolio Project – TELANGANA GROWTH ANALYSIS

Platform Used for Data Collection:

 Kaggle for data collection: https://www.kaggle.com/datasets/mohamedharris/restaurant-orderdetails/data

Software Used:

- 1) MS Excel for storing the data and visualization of the results
- 2) My SQL for performing the analysis by writing queries





PROCESS FOLLOWED



- 1) Downloading dataset from Kaggle in CSV Format.
- 2) Create a Database and using it for storing the tables using Create and Use command.
- 3) Importing Data from CSV into MySQL Database by using table import wizard of MySQL.



- 1) Checking the datatypes of all the tables.
- 2) Preparation of the data for analysis by converting the datatype of certain columns of a specific table as per requirement using alter commands.



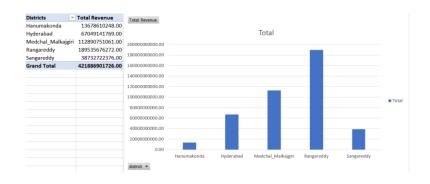
- 1) Utilized **SQL** to extract data by writing queries from different related tables from the databases using **JOIN**, **Subqueries**, **CASE statement**.
- 2) Transformed and filtered data by using aggregating and filtering function, window function, date functions to improve reporting process.
- 3) Loaded and visualized data with **Excel** to identify key factors the growth of the Telangana State.

Stamp Registration

1. List down the Top 5 districts with highest and lowest total revenue generated from documents and stamps between FY 2019 and FY 2022?

Top 5 districts with highest total revenue generated from documents and stamps between FY 2019 and FY 2022

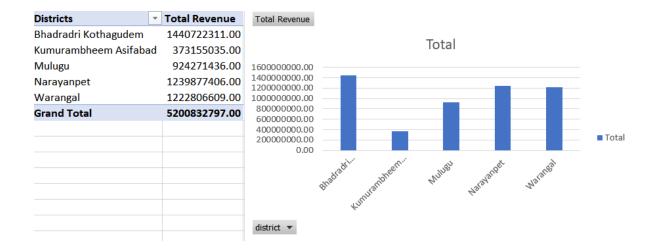
```
select d.district ,sum(f.documents_registered_rev+ f.estamps_challans_rev) as total_revenue
from fact_stamps f
join dim_districts d
on d.dist_code = f.dist_code
join dim_date dd
on dd.month = f.month
where dd.fiscal_year between 2019 and 2022
group by d.district
order by total_revenue desc
limit 5;
                 total_revenue
 district
 Rangareddy
                     189535676272
 Medchal_Malkajgiri 112890751061
 Hyderabad
                    67049141769
 Sangareddy
                    38732722376
 Hanumakonda
                    13678610248
```



Top 5 districts with lowest total revenue generated from documents and stamps between FY 2019 and FY 2022

```
select d.district ,sum(f.documents_registered_rev+ f.estamps_challans_rev) as total_revenue
from fact_stamps f
join dim_districts d
on d.dist_code = f.dist_code
join dim_date dd
on dd.month = f.month
where dd.fiscal_year between 2019 and 2022
group by d.district
order by total_revenue asc
limit 5;
```

	district	total_revenue
١	Kumurambheem Asifabad	373155035
	Mulugu	924271436
	Warangal	1222806609
	Narayanpet	1239877406
	Bhadradri Kothagudem	1440722311

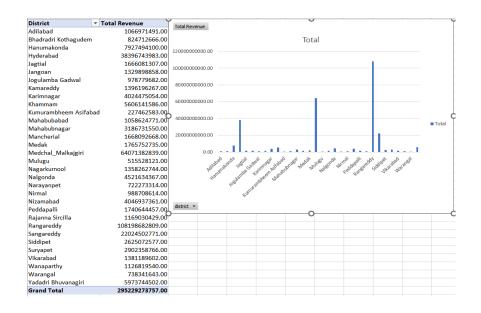


2. How does the revenue generated from document registration vary across districts in Telangana?

```
select distinct(dim_districts.district), sum(fact_stamps.documents_registered_rev) as Total_Revenue
from fact_stamps
join dim_districts on fact_stamps.dist_code = dim_districts.dist_code
group by dim_districts.district
order by Total_Revenue desc;
```

	district	Total_Revenue	
•	Rangareddy	108198682809	
	Medchal_Malkajgiri	64071382839	
	Hyderabad	38396743983	
	Sangareddy	22024502771	
	Hanumakonda	7927494100	
	Yadadri Bhuvanagiri	5973744502	
	Khammam	5606141586	
	Nalgonda	4521634367	
	Nizamabad	4046937361	
	Karimnagar	4024475054	
	Mahabubnagar	3186731550	
	Suryapet	2902358766	
	Siddipet	2625072577	
	Medak	1765752735	
	Peddapalli	1740644457	
	Mancherial	1668092668	
	Jagtial	1666081307	
	Kamareddy	1396196267	
	Vikarabad	1381189602	
	Nagarkurnool	1358262744	
	Jangoan	1329898858	
	Rajanna Sircilla	1169030429	
	Wanaparthy	1126819540	
	Adilabad	1066971491	
	Mahabubabad	1058624771	

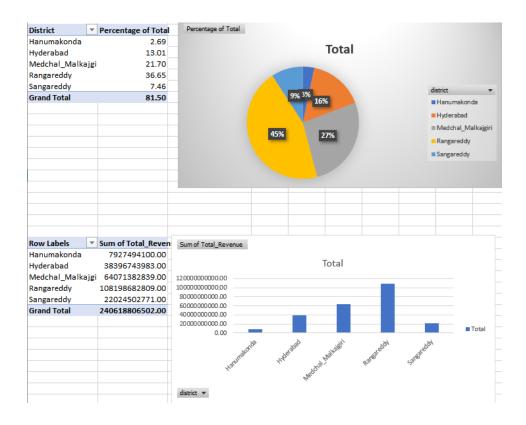
988708614
978779682
824712666
738341643
722273314
515528121
227462583



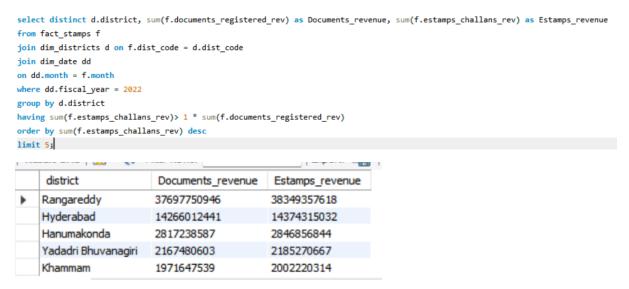
3. List down the top 5 districts that showed the highest document registration revenue growth between FY 2019 and 2022.

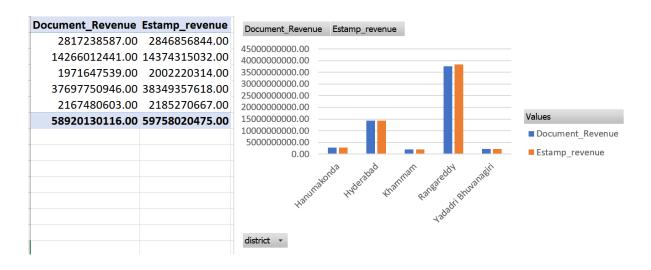
```
select distinct(d.district), sum(f.documents_registered_rev) as Total_Revenue, (sum(f.documents_registered_rev)*100/(select sum(documents_registered_rev) from fact_stamps)) as percentage_total from fact_stamps f
join dim_districts d on f.dist_code = d.dist_code
join dim_date dd
on dd.month = f.month
where dd.fiscal_year between 2019 and 2022
group by d.district
order by Total_Revenue desc
limit 5;
```

	district	Total_Revenue	percentage_total
•	Rangareddy	108198682809	36.6490
	Medchal_Malkajgiri	64071382839	21.7022
	Hyderabad	38396743983	13.0057
	Sangareddy	22024502771	7.4601
	Hanumakonda	7927494100	2.6852

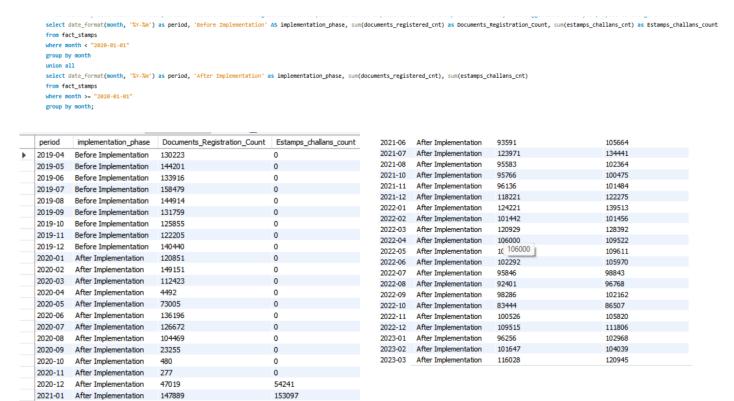


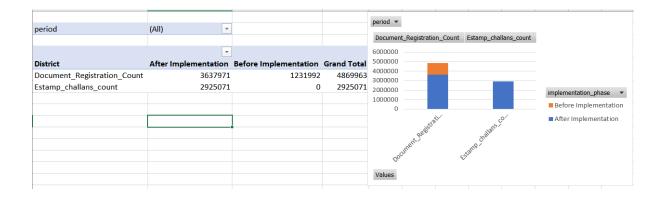
4. How does the revenue generated from document registration compare to the revenue generated from e-stamp challans across districts? List down the top 5 districts where e-stamps revenue contributes significantly more to the revenue than the documents in FY 2022?





5. Is there any alteration of e-Stamp challan count and document registration count pattern since the implementation of e-Stamp challan?





140062

148732

111412

2021-02

2021-04

After Implementation

After Implementation

2021-03 After Implementation

2021-05 After Implementation

137407

142732

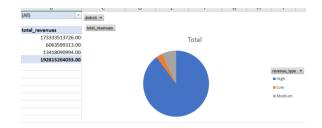
107210

6. Categorize districts into three segments based on their stamp registration revenue generation during the fiscal year 2021 to 2022.

```
select dim_districts.district, sum(fact_stamps.estamps_challans_rev) as total_revenue,
case
when ntile(3) over (order by sum(fact_stamps.estamps_challans_rev)) = 1 then 'Low'
when ntile(3) over (order by sum(fact_stamps.estamps_challans_rev)) = 2 then 'Medium'
when ntile(3) over (order by sum(fact_stamps.estamps_challans_rev)) = 3 then 'High'
end as revenue_type
from fact_stamps
join dim_districts
on fact_stamps.dist_code = dim_districts.dist_code
join dim_date
on fact_stamps.month = dim_date.month
where (dim_date.fiscal_year) between "2021" and "2022"
group by dim_districts.district;
```

	district	total_revenue	revenue_type
٠	Kumurambheem Asifabad	133420327	Low
	Mulugu	369302299	Low
	Warangal	424814165	Low
	Narayanpet	468512790	Low
	Bhadradri Kothagudem	540722552	Low
	Jogulamba Gadwal	603996030	Low
	Nirmal	639557126	Low
	Mahabubabad	666572041	Low
	Adilabad	723003081	Low
	Wanaparthy	746269937	Low
	Jangoan	747428965	Low
	Rajanna Sircilla	750209633	Medium
	Kamareddy	859661958	Medium
	Nagarkurnool	881618696	Medium
	Vikarabad	906189501	Medium
	Jagtial	1079332319	Medium
	Mancherial	1105983406	Medium
	Peddapalli	1126687326	Medium
	Medak	1146081195	Medium
	Siddipet	1762872744	Medium
	Suryapet	1818082431	Medium
	Mahabubnagar	1981371785	Medium
	Nalgonda	2600080824	High
	Nizamabad	2614662871	High
	Karimnagar	2633977624	High
	Khammam	3646844602	High

		_
Yadadri Bhuvanagiri	3801870232	High
Hanumakonda	4959670232	High
Sangareddy	14782041561	High
Hyderabad	25267992981	High
Medchal_Malkajgiri	42344075412	High
Rangareddy	70682297387	High
Rangareddy	70682297387	High



Transportation

7. Investigate whether there is any correlation between vehicle sales and specific months or seasons in different districts. Are there any months or seasons that consistently show higher or lower sales rate, and if yes, what could be the driving factors? (Consider Fuel-Type category only)

select distinct monthname(month) as month, sum(fuel_type_petrol+ fuel_type_diesel+fuel_type_electric+fuel_type_others) as vehicle_sales
from fact_transport
group by monthname(month)
order by vehicle_sales desc;

	month	vehide_sales	
•	October	748812	
	June	613414	
	March	600458	
	November	586376	
	August	569051	
	January	564491	
	July	558992	
	February	535443	
	September	510831	
	December	507991	
	May	439195	
	April	430723	



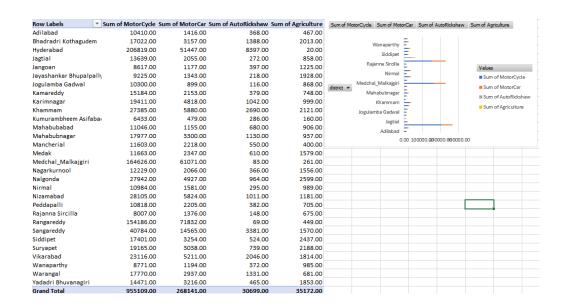
8. How does the distribution of vehicles vary by vehicle class (MotorCycle, MotorCar, AutoRickshaw, Agriculture) across different districts? Are there any districts with a predominant preference for a specific vehicle class? Consider FY 2022 for analysis.

select distinct(d.district), sum(t.vehicleClass_MotorCycle) as MotorCycle, sum(t.vehicleClass_MotorCar) as MotorCar, sum(t.vehicleClass_AutoRickshaw) as AutoRickshaw, sum(t.vehicleClass_Agriculture) as Agriculture from fact_transport t
join dim_districts d
on t.dist_code = d.dist_code
join dim_date dd
on t.month = dd.month
where dd.fiscal_year = 2022
group by d.district

	district	MotorCycle	MotorCar	AutoRickshaw	Agriculture
•	Hyderabad	206819	51447	8397	20
	Medchal_Malkajgiri	164626	61071	83	261
	Rangareddy	154186	71832	69	449
	Sangareddy	40784	14565	3381	1570
	Nizamabad	28105	5824	1011	1181
	Nalgonda	27942	4927	964	2599
	Khammam	27385	5880	2690	2121
	Vikarabad	23116	5211	2046	1814
	Karimnagar	19411	4818	1042	999
	Suryapet	19165	3038	739	2188
	Mahabubnagar	17977	3300	1130	937
	Warangal	17770	2937	1331	681
	Siddipet	17401	3254	524	2437
	Bhadradri Kothag	17022	3157	1388	2013
	Kamareddy	15184	2153	379	748
	Yadadri Bhuvanagiri	14471	3216	465	1853
	Jagtial	13639	2055	272	858
	Nagarkurnool	12229	2066	366	1556
	Medak	11663	2347	610	1579
	Mancherial	11603	2218	550	400
	Mahabubabad	11046	1155	680	906
	Nirmal	10984	1581	295	989
	Peddapalli	10818	2205	382	705
	Adilabad	10410	1416	368	467
	Jogulamba Gadwal	10300	899	116	868
	Jayashankar Bhu	9225	1343	218	1928

order by sum(t.vehicleClass_MotorCycle) desc;

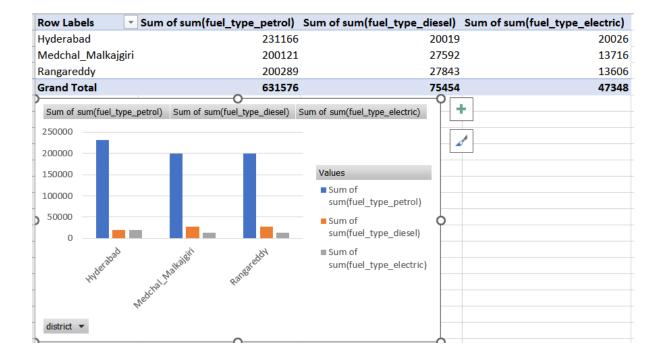
,				
Wanaparthy	8771	1194	372	985
Jangoan	8617	1177	397	1225
Rajanna Sircilla	8007	1376	148	675
Kumurambheem A	6433	479	286	160



9. List down the top 3 and bottom 3 districts that have shown the highest and lowest vehicle sales growth during FY 2022 (Consider and compare categories: Petrol, Diesel and Electric)

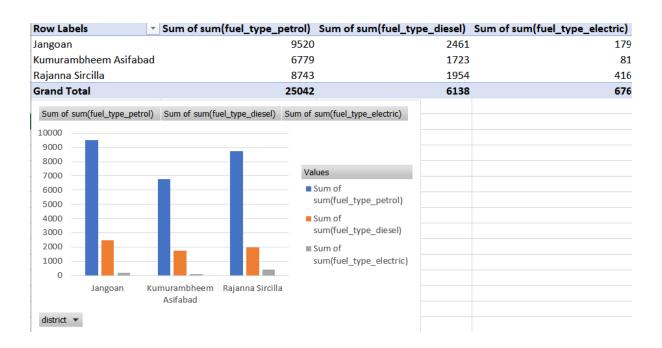
```
-- Top 3 district with highest sales in 2022
select distinct d.district,sum(fuel_type_petrol), sum(fuel_type_diesel), sum(fuel_type_electric)
from fact_transport t
join dim_districts d
on t.dist_code = d.dist_code
join dim_date dd
on t.month = dd.month
where dd.fiscal_year = '2022'
group by d.district
order by sum(fuel_type_diesel) desc
limit 3;
```

	district	sum(fuel_type_petrol)	sum(fuel_type_diesel)	sum(fuel_type_electric)
•	Rangareddy	200289	27843	13606
	Medchal_Malkajgiri	200121	27592	13716
	Hyderabad	231166	20019	20026



```
-- Top 3 district with lowest sales in 2022
select distinct d.district,sum(fuel_type_petrol), sum(fuel_type_diesel), sum(fuel_type_electric)
from fact_transport t
join dim_districts d
on t.dist_code = d.dist_code
join dim_date dd
on t.month = dd.month
where dd.fiscal_year = '2022'
group by d.district
order by sum(fuel_type_petrol) asc
limit 3;
```

	district	sum(fuel_type_petrol)	sum(fuel_type_diesel)	sum(fuel_type_electric)
•	Kumurambheem Asifabad	6779	1723	81
	Rajanna Sircilla	8743	1954	416
	Jangoan	9520	2461	179

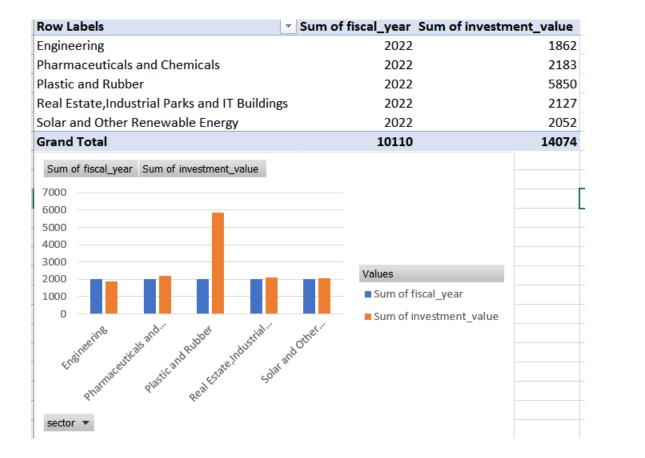


Telangana State Industrial Project Approval and Self Certification System

10. List down the top 5 sectors that have witnessed the most significant investments in FY 2022.

```
select (d.fiscal_year),f.sector, sum(f.`investment in cr`) as investment_value
from fact_ts_ipass f
join dim_date d
on f.month = d.month
where (d.fiscal_year) = 2022
group by d.fiscal_year, f.sector
order by sum(f.`investment in cr`) desc
limit 5;
```

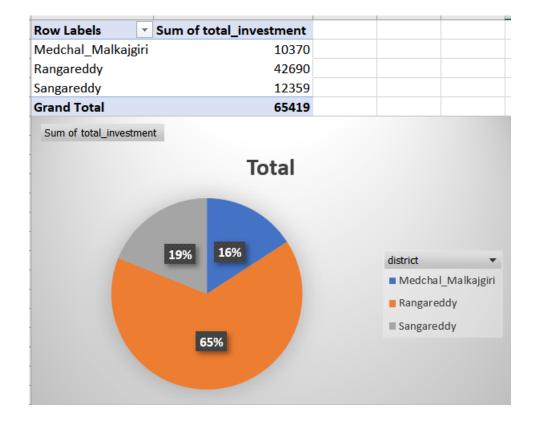
	fiscal_year	sector	investment_value
٠	2022	Plastic and Rubber	5850
	2022	Pharmaceuticals and Chemicals	2183
	2022	Real Estate, Industrial Parks and IT Buildings	2127
	2022	Solar and Other Renewable Energy	2052
	2022	Engineering	1862



11.List down the top 3 districts that have attracted the most significant sector investments during FY 2019 to 2022?

```
select ds.district, sum(f.`investment in cr`) as total_investment
from fact_ts_ipass f
join dim_date d
on f.month = d.month
join dim_districts ds
on ds.dist_code = f.dist_code
where (d.fiscal_year) between '2019' and '2022'
group by ds.district
order by sum(f.`investment in cr`) desc
limit 3;
```

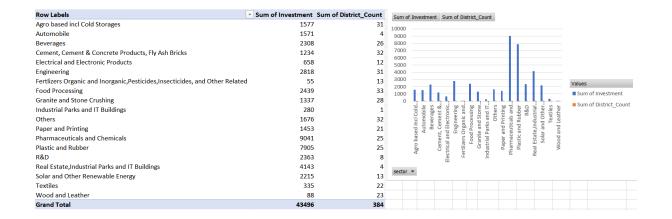
	district	total_investment
١	Rangareddy	42690
	Sangareddy	12359
	Medchal_Malkajgiri	10370



12. Are there any particular sectors that have shown substantial investment in multiple districts between FY 2021 and 2022?

```
select fi.sector, count(distinct d.district) as District_Count, sum(fi.`investment in cr`) as Investment
from fact_ts_ipass fi
join dim_districts d
on d.dist_code = fi.dist_code
join dim_date dd
on dd.month = fi.month
where dd.fiscal_year between 2021 and 2022
group by fi.sector
order by count(d.district) desc;
```

	sector	District_Count	Investment
•	Food Processing	33	2439
	Engineering	31	2818
	Cement, Cement & Concrete Products, Fly Ash	32	1234
	Agro based ind Cold Storages	31	1577
	Others	32	1676
	Granite and Stone Crushing	28	1337
	Pharmaceuticals and Chemicals	25	9041
	Plastic and Rubber	25	7905
	Beverages	26	2308
	Wood and Leather	23	88
	Paper and Printing	21	1453
	Textiles	22	335
	Electrical and Electronic Products	12	658
	R&D	8	2363
	Fertlizers Organic and Inorganic, Pesticides, Inse	13	55
	Solar and Other Renewable Energy	13	2215
	Automobile	4	1571
	Real Estate, Industrial Parks and IT Buildings	4	4143
	Industrial Parks and IT Buildings	1	280



13. Can we identify any seasonal patterns or cyclicality in the investment trends for specific sectors? Do certain sectors experience higher investments during particular months?

```
select dd.Mmm, fi.sector, sum(fi.`investment in cr`) as total_investment
from fact_ts_ipass fi
join dim_date dd
on dd.month = fi.month
group by dd.Mmm, fi.sector
order by total_investment desc;
```

	sector	Mmm	Total_Investment
•	Real Estate, Industrial Parks and IT Buildings	Feb	17843
	Fertlizers Organic and Inorganic, Pesticides, Inse	Dec	5274
	Real Estate, Industrial Parks and IT Buildings	Mar	3399
	Pharmaceuticals and Chemicals	Jul	2807
	Pharmaceuticals and Chemicals	Sep	2140
	Pharmaceuticals and Chemicals	Dec	2107
	Real Estate, Industrial Parks and IT Buildings	Aug	2008
	Solar and Other Renewable Energy	Nov	2008
	Plastic and Rubber	Dec	1956
	Pharmaceuticals and Chemicals	Feb	1858
	Plastic and Rubber	Jan	1800
	Pharmaceuticals and Chemicals	Jun	1694
	Plastic and Rubber	Aug	1568
	Real Estate, Industrial Parks and IT Buildings	Jan	1397
	Engineering	Feb	1330
	Automobile	Jan	1254
	Solar and Other Renewable Energy	Apr	1203
	Textiles	Sep	1191
	Real Estate, Industrial Parks and IT Buildings	Jul	1124
	Solar and Other Renewable Energy	Jun	1109
	Plastic and Rubber	Apr	1093
	Paper and Printing	Jul	1060

sector	Mmm	Total_Investment
Wood and Leather	Feb	9
Fertlizers Organic and Inorganic,Pesticides,Inse	Apr	8
Fertlizers Organic and Inorganic,Pesticides,Inse	Jun	8
Wood and Leather	Aug	8
Beverages	Sep	7
Wood and Leather	Dec	7
Automobile	Mar	7
Wood and Leather	Apr	6
Fertlizers Organic and Inorganic,Pesticides,Inse	Jul	5
Fertlizers Organic and Inorganic,Pesticides,Inse	May	4
Fertlizers Organic and Inorganic,Pesticides,Inse	Nov	4
Wood and Leather	Mar	4
Solar and Other Renewable Energy	Feb	4
Wood and Leather	May	3
Automobile	Apr	2
Beverages	Aug	1
Wood and Leather	Oct	1
Solar and Other Renewable Energy	Oct	1
Automobile	Dec	1
Automobile	Nov	0
Automobile	Feb	0
Automobile	Jun	0

