

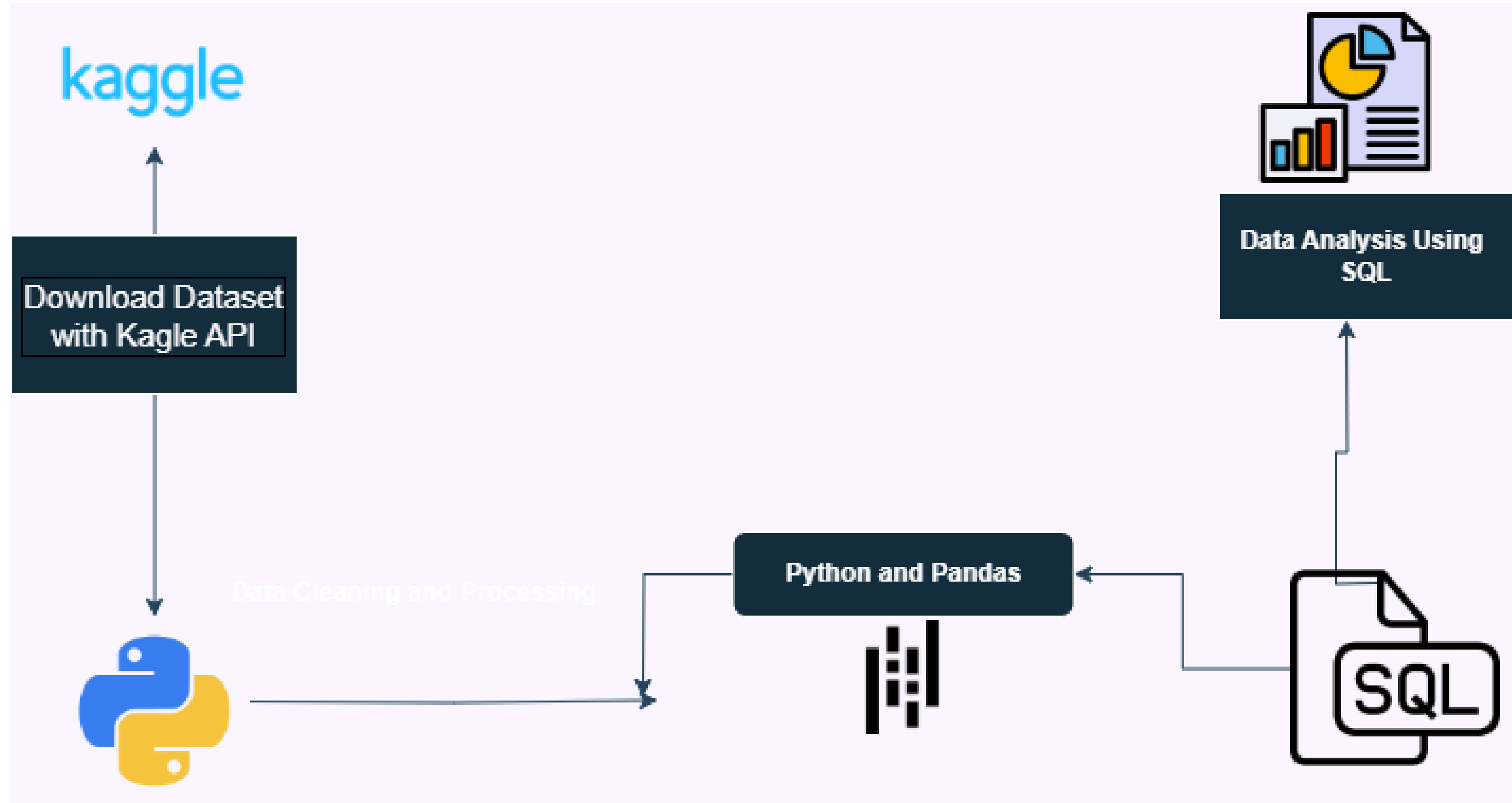
kaggle



Data Analytics with Python & SQL



Project Architecture



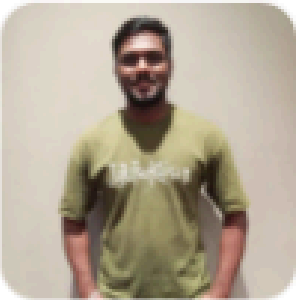
SQL Project Questions


1. find the 10 highest revenue base product
2. find top 5 highest selling product in each region
3. find month over month growth comparison for 2022 and 2023 sales eg: jan 2022 vs jan 2023
4. for each category which month had highest sales
5. which sub category had highest growth by profit in 2023 compare to 2022


Data Cleaning and Processing with Python


Check it out link:


parchakeavinash/**Data-Analytics-with-SQL-and...**




 1
Contributor

 0
Issues


 0
Stars

 0
Forks



Data-Analytics-with-SQL-and-Python/SQL Project.ipynb at main · parchakeavinash/Data-Analytics-with-SQL-and-...

Contribute to parchakeavinash/Data-Analytics-with-SQL-and-Python development by creating an account on GitHub.

 GitHub

1.find the 10 highest revenue base product

```
• SELECT
    product_id, SUM(sales_price) AS sales
FROM
    orders
GROUP BY product_id
ORDER BY sales DESC
LIMIT 10;
```

Output:

Result Grid			Filter Rows:	Export:	Write
	product_id	sales			
▶	TEC-CO-10004722	59514.00			
	OFF-BI-10003527	26525.30			
	TEC-MA-10002412	21734.40			
	FUR-CH-10002024	21096.20			
	OFF-BI-10001359	19090.20			
	OFF-BI-10000545	18249.00			
	TEC-CO-10001449	18151.20			
	TEC-MA-10001127	17906.40			

Result 10 ×

2. find top 5 highest selling product in each region

```
• SELECT
    region,
    product_id,
    sales
FROM
    (SELECT
        region,
        product_id,
        SUM(sales_price) AS sales,
        ROW_NUMBER() OVER (PARTITION BY region ORDER BY SUM(sales_price) DESC) AS sales_rank
    FROM
        orders
    GROUP BY
        region, product_id
    ) AS rankproduct
WHERE
    sales_rank <= 5;
```

2. find top 5 highest selling product in each region

Output:

	region	product_id	sales
▶	Central	TEC-CO-10004722	16975.00
	Central	TEC-MA-10000822	13770.00
	Central	OFF-BI-10001120	11056.50
	Central	OFF-BI-10000545	10132.70
	Central	OFF-BI-10004995	8416.10
	East	TEC-CO-10004722	29099.00
	East	TEC-MA-10001047	13767.00
	East	FUR-BO-10004834	11274.10
	East	OFF-BI-10001359	8463.60
	East	TEC-CO-10001449	8316.00
	South	TEC-MA-10002412	21734.40
	South	TEC-MA-10001127	11116.40
	South	OFF-BI-10001359	8053.20
	South	TEC-MA-10004135	7848.00

3. find month over month growth comparison for 2022 and 2023 sales eg: jan 2022 vs jan 2023

```
with cte as (  
    select year(order_date) as order_year, month(order_date) as order_month, sum(sales_price) as sales  
    from orders  
    group by order_year, order_month  
    -- order by order_year, order_month  
)  
select order_month  
    , sum(case when order_year = 2022 then sales else 0 end) as 2022_sales  
    , sum(case when order_year = 2023 then sales else 0 end) as 2023_sales  
from cte  
group by order_month  
order by order_month;
```


Result Grid



Filter Rows:

Export:



Wrap Cell Content:



order_month

2022_sales

2023_sales

1

94712.50

88632.60

2

90091.00

128124.20

3

80106.00

82512.30

4

95451.60

111568.60

5

79448.30

86447.90

6

94170.50

68976.50

7

78652.20

90563.80

8

104808.00

87733.60

9

79142.20

76658.60

10

118912.70

121061.50

11

84225.30

75432.80

Result 24



4.for each category which month had highest sales

```
107 • with cte as (  
108     select category,DATE_FORMAT(order_date, '%Y%m') AS order_year_month, sum(sales_price) as sales  
109     from orders  
110     group by category,DATE_FORMAT(order_date, '%Y%m')  
111     order by category,DATE_FORMAT(order_date, '%Y%m')  
112 )  
113 • select * from(  
114     select *,  
115     row_number() over(partition by category order by sales desc) as rn  
116     from cte) as a  
117 where rn =1;
```

Output:

Result Grid Filter Rows: Export: Wrap Cell Content:				
	category	order_year_month	sales	rn
▶	Furniture	202210	42888.90	1
	Office Supplies	202302	44118.50	1
	Technology	202310	53000.10	1

5. which sub category had highest growth by profit in 2023 compare to 2022

```
• with cte as (  
  select sub_category, year(order_date) as order_year, sum(sales_price) as sales  
  from orders  
  group by sub_category, order_year  
  -- order by order_year, order_month  
  )  
  , cte2 as(  
  select sub_category  
  , sum(case when order_year = 2022 then sales else 0 end) as 2022_sales  
  , sum(case when order_year = 2023 then sales else 0 end) as 2023_sales  
  from cte  
  group by sub_category  
  order by sub_category  
  )  
  select *  
  , (2023_sales - 2022_sales)*100/2022_sales as "%growth"  
  from cte2  
  order by (2023_sales - 2022_sales)*100/2022_sales desc;  
  -- limit 1;
```

5. which sub category had highest growth by profit in 2023 compare to 2022

Output:

Result Grid Filter Rows: Export: Wrap Cell Content:				
	sub_category	2022_sales	2023_sales	%growth
▶	Supplies	16140.70	28917.40	79.158277
	Machines	73723.20	109178.50	48.092459
	Binders	87675.50	108363.10	23.595645
	Storage	102907.40	113000.60	9.808041
	Chairs	151395.30	165429.80	9.270103
	Accessories	77627.20	83977.40	8.180380
	Bookcases	53469.50	57346.60	7.251050
	Fasteners	1430.10	1508.80	5.503112
	Phones	157334.70	160673.60	2.122164
	Paper	38898.90	36932.40	-5.055413
	Art	13644.10	12615.70	-7.537324
	Labels	6329.60	5665.30	-10.495134
	Furnishings	47816.20	40522.10	-15.254454

Result 31 ×

Thank You!!!