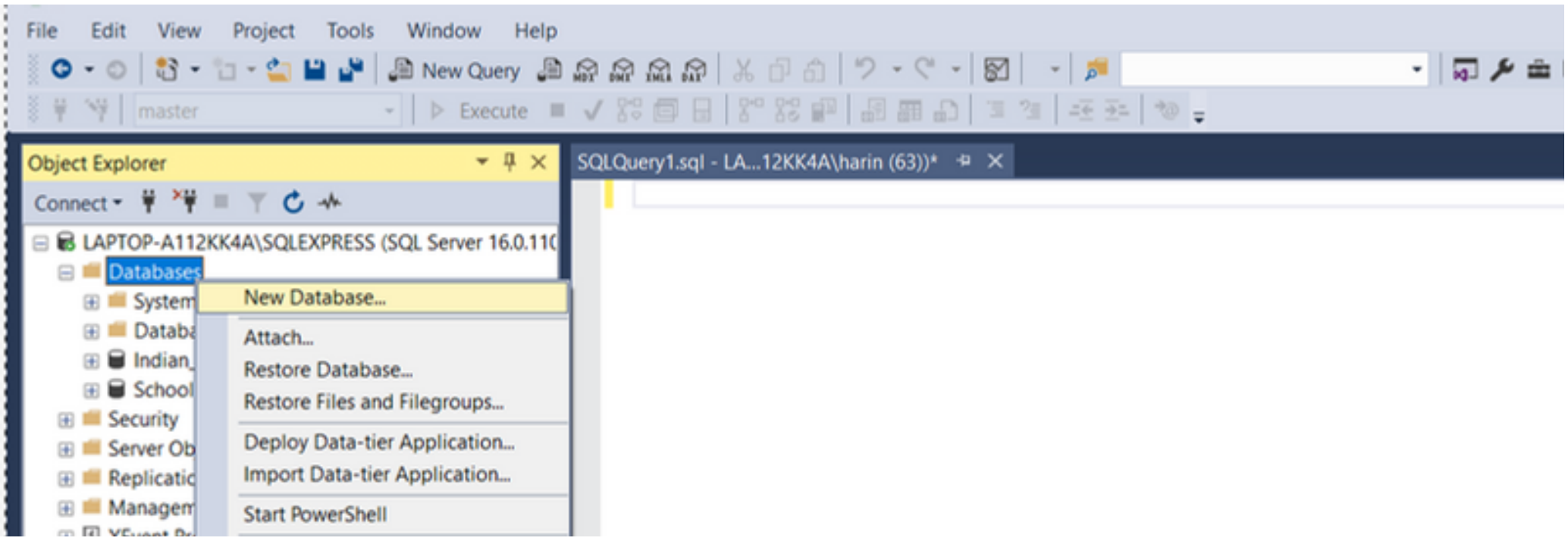
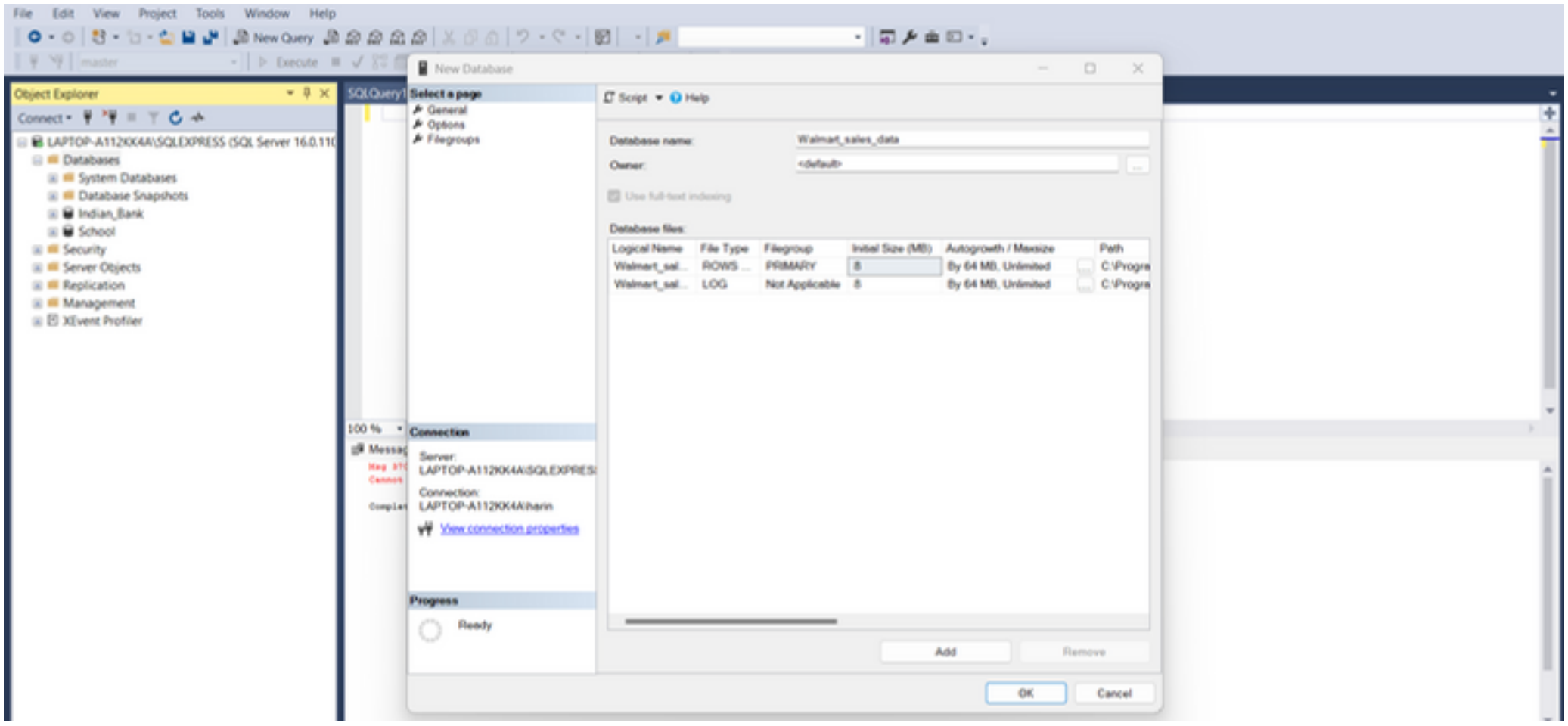
Analysis of Walmart Sales Data using SQL

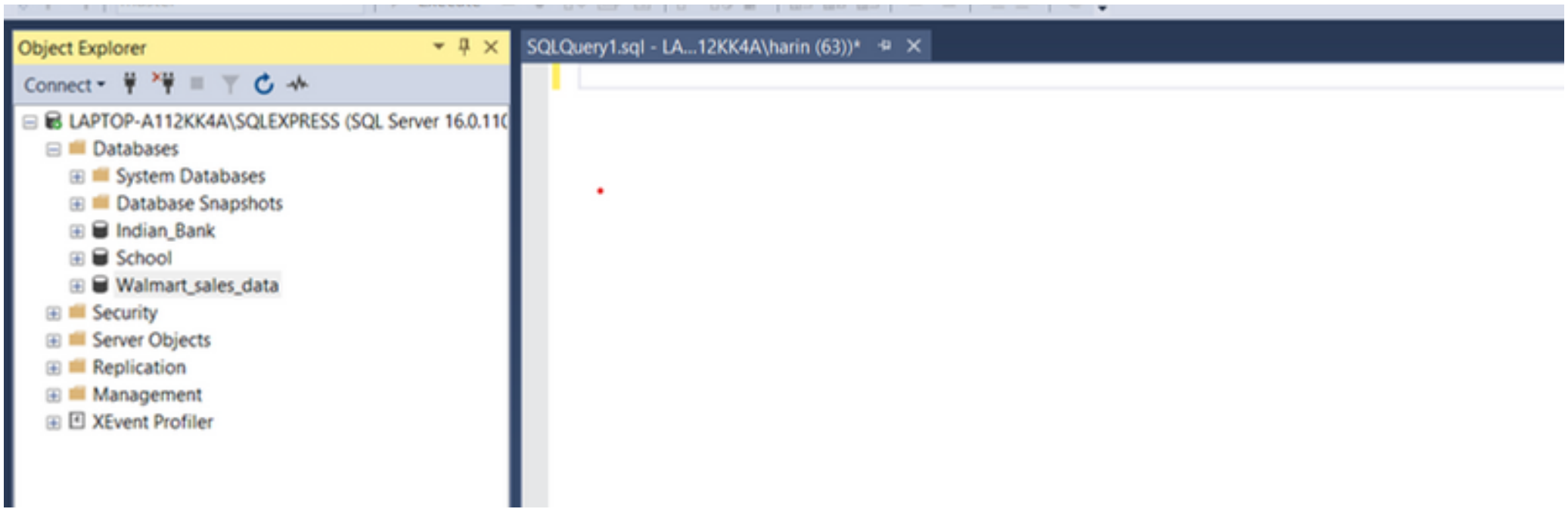
Harikrishnan Nair

Index

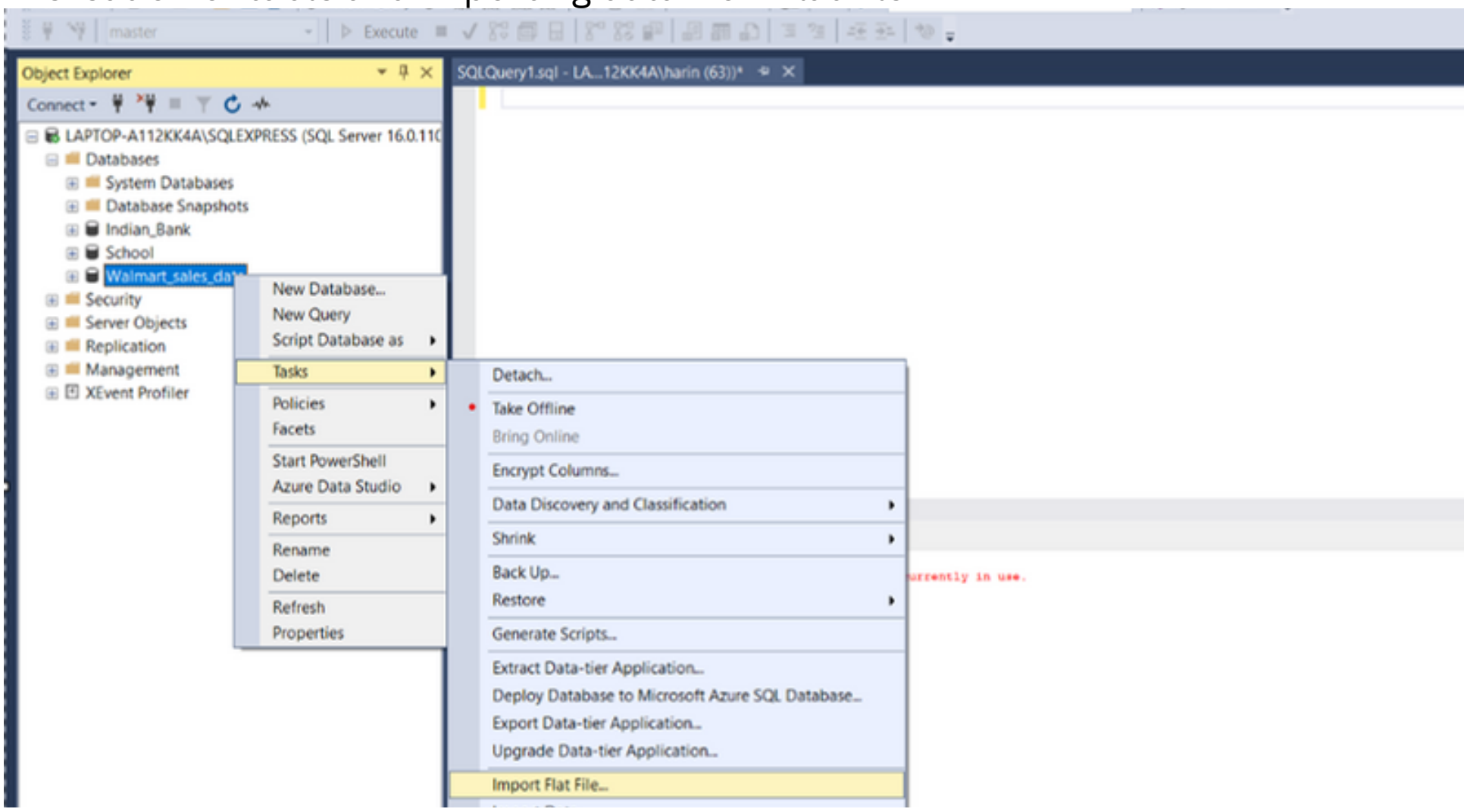
- Creation of database and table of Walmart sales data

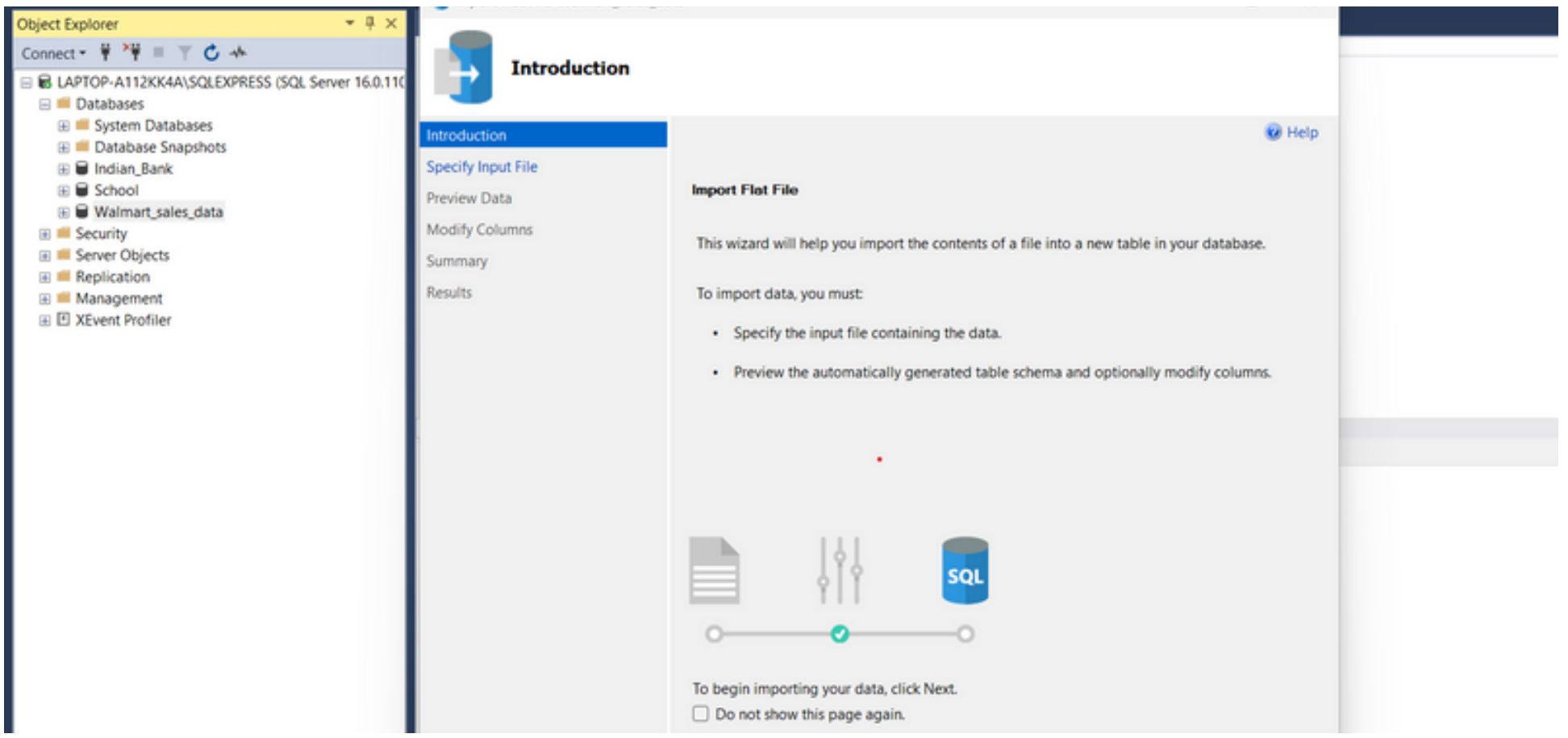




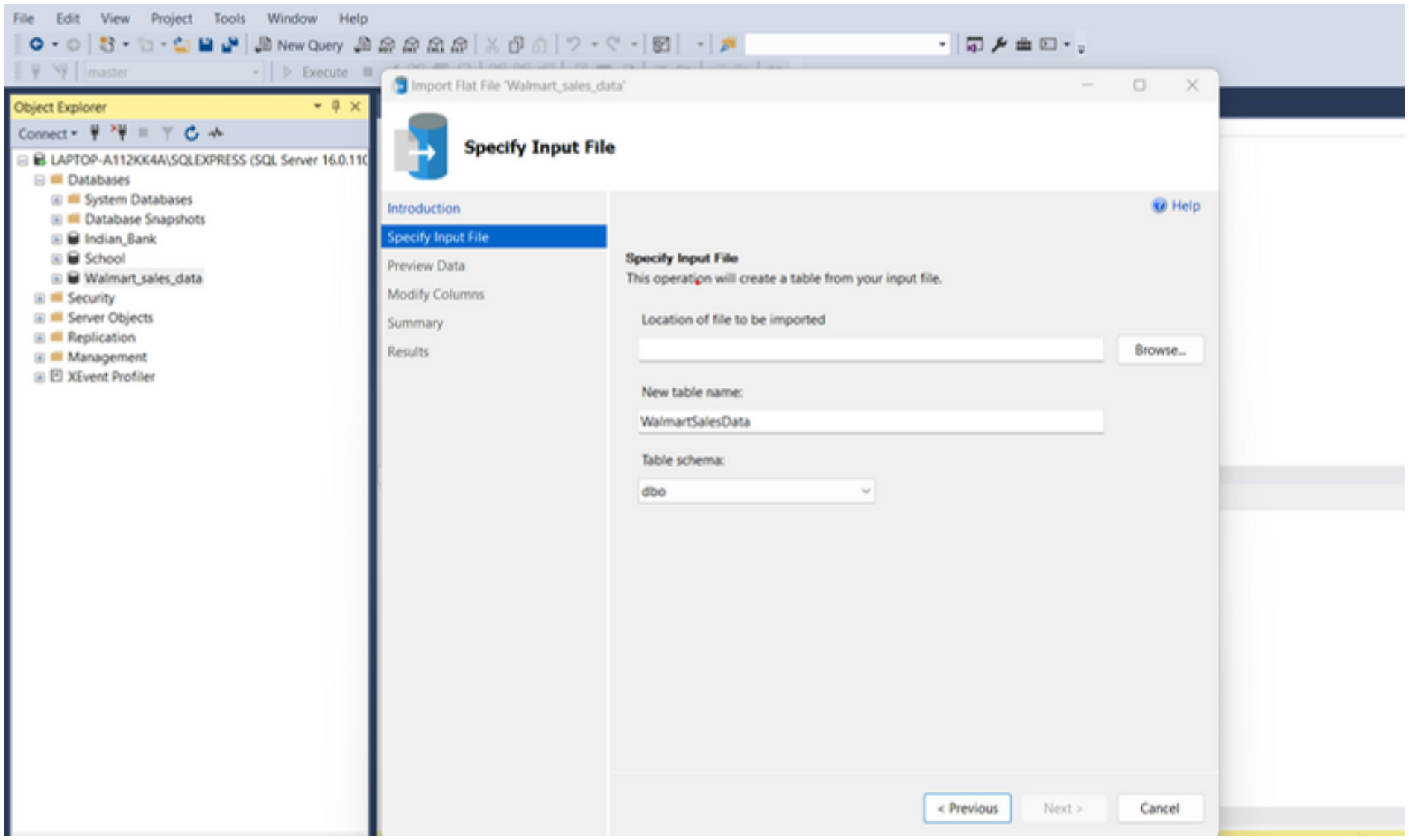


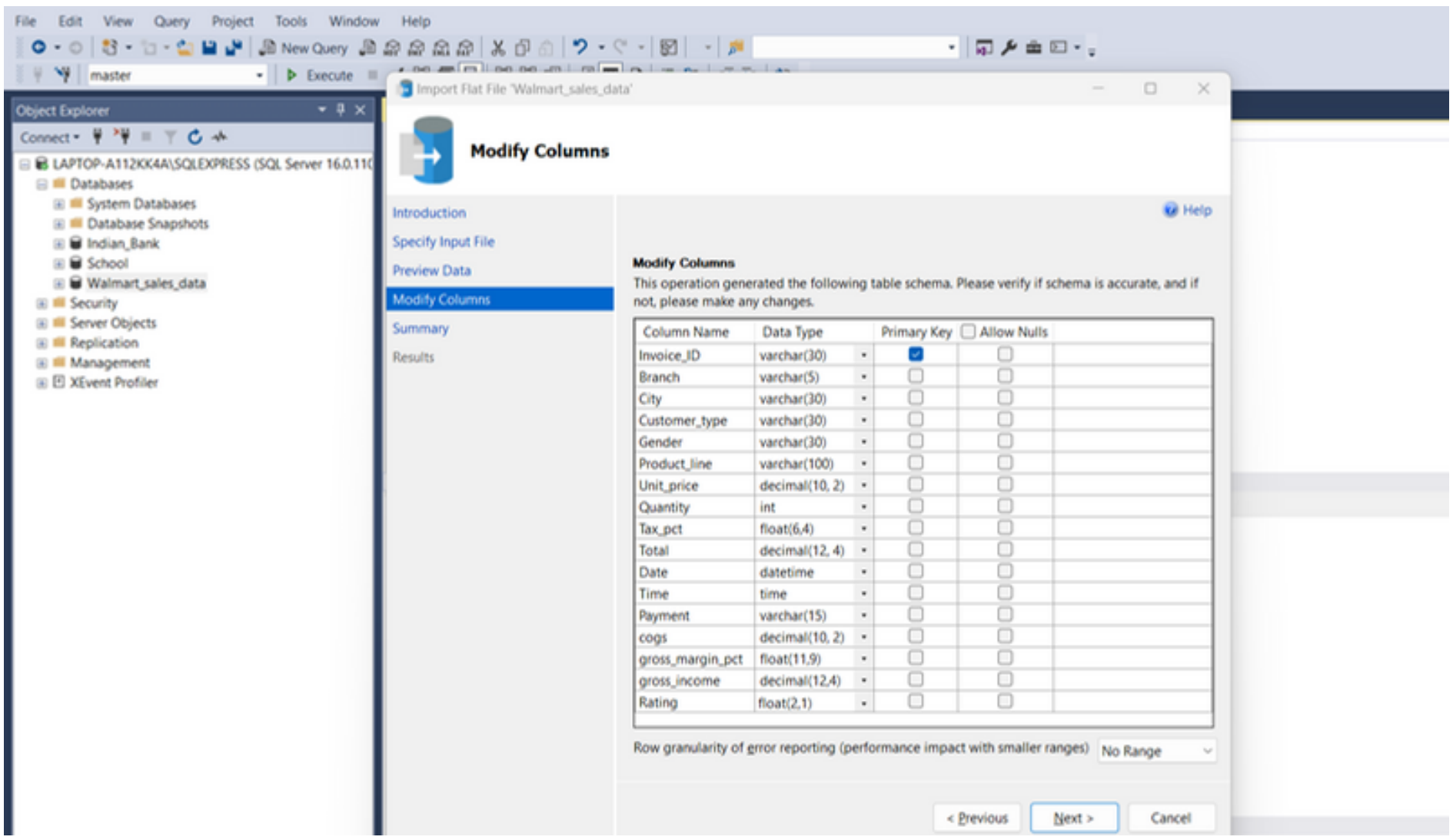
- Creation of table and importing data from flat file



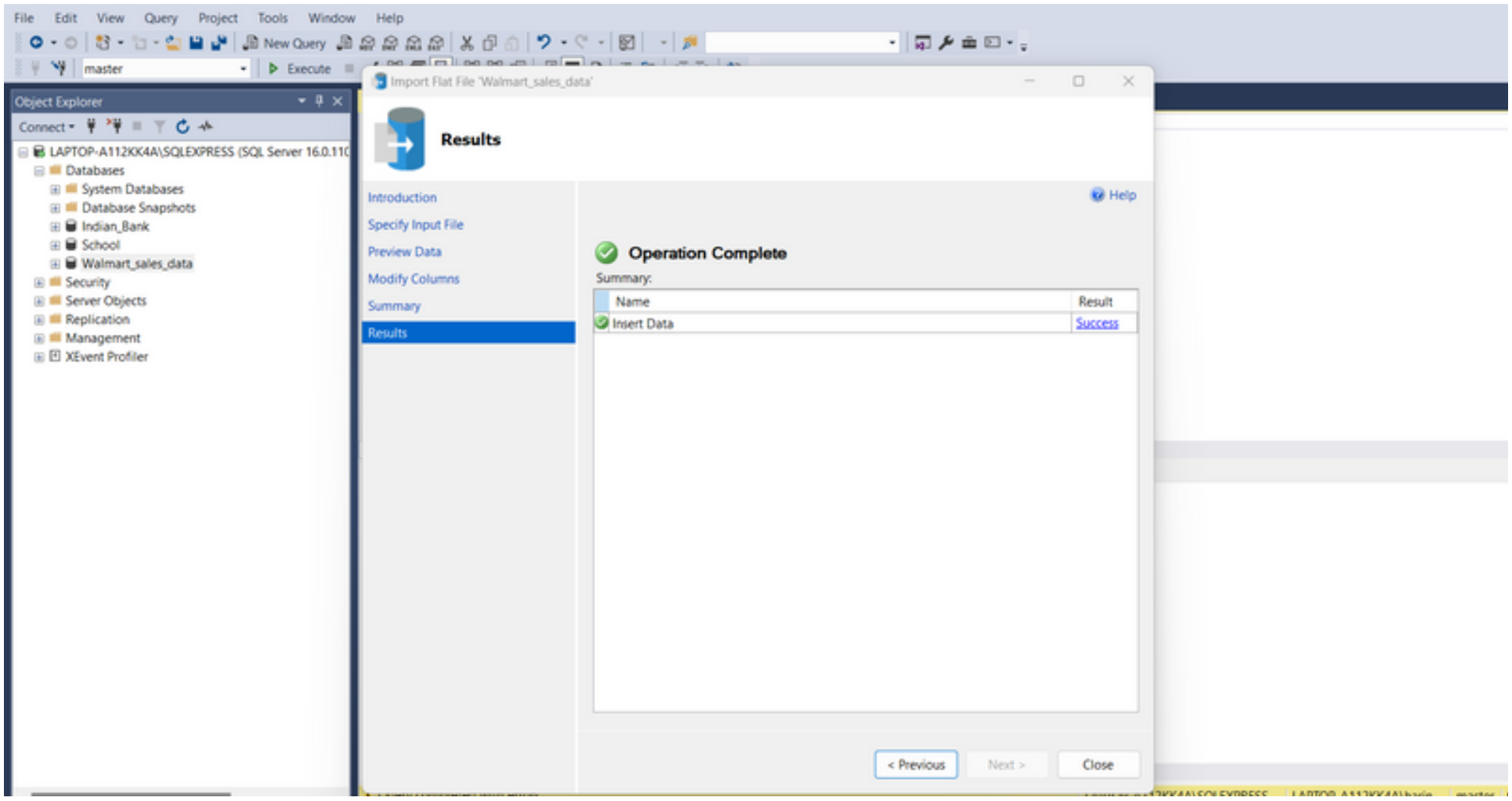


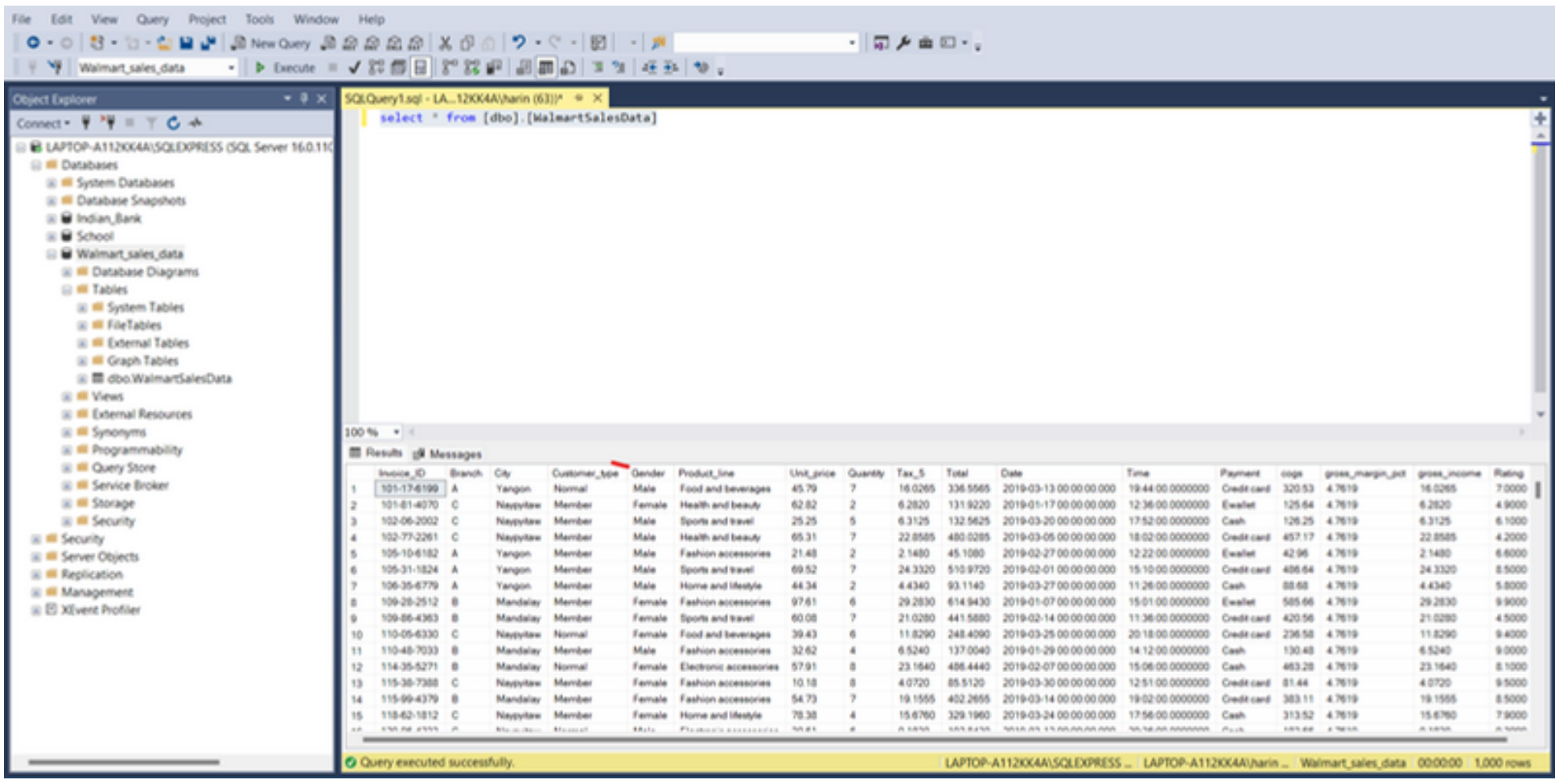
- Defining columns and datatypes





- Finishing the data import process and checking the imported data by querying





- **Data Analysis based on Generic queries**

1. How many unique cities does the data have?

```
select distinct(city) as city_name
from WalmartSalesData
```

Output:

	city_name
1	Mandalay
2	Yangon
3	Naypyitaw

2. In which city is each branch ?

```
select distinct(city), Branch
from WalmartSalesData
```

Output:

	city	Branch
1	Naypyitaw	C
2	Yangon	A
3	Mandalay	B

- **Data Analysis of Product-related queries**

1. How many unique lines does the data have?

```
select distinct(product_line)
from WalmartSalesData
```

Output:

	product_line
1	Fashion accessories
2	Home and lifestyle
3	Electronic accessories
4	Health and beauty
5	Food and beverages
6	Sports and travel

2. What is the most common payment method?

```
select Payment, count(Payment) Payment_Cnt from WalmartSalesData
group by Payment
order by Payment_Cnt desc
```

Output:

	Payment	Payment_Cnt
1	Ewallet	345
2	Cash	344
3	Credit card	311

3. What is the most selling product line?

```
select Product_line, count(Product_line) Product_line_cnt from WalmartSalesData
group by Product_line
order by Product_line_cnt  desc
```

Output:

	Product_line	Product_line_cnt
1	Fashion accessories	178
2	Food and beverages	174
3	Electronic accessories	170
4	Sports and travel	166
5	Home and lifestyle	160
6	Health and beauty	152

4. What is the total revenue by month?

```
select Month_name, sum(Total) Total_revenue from WalmartSalesData
group by Month_name
order by Total_revenue desc
```

Output:

	Month_name	Total_revenue
1	January	116291.8680
2	March	109455.5070
3	February	97219.3740

5. What month had the largest Cost of Goods Sold(COGS)?

```
select Month_name, sum(cogs) Cost_of_Goods_sold_COGS from WalmartSalesData
group by Month_name
order by Cost_of_Goods_sold_COGS desc
```

Output:

	Month_name	Cost_of_Goods_sold_COGS
1	January	110754.16
2	March	104243.34
3	February	92589.88

6. What product line had the largest revenue?

```
select Product_line, sum(total) Total_revenue from WalmartSalesData
group by Product_line
order by Total_revenue desc
```

Output:

	Product_line	Total_revenue
1	Food and beverages	56144.8440
2	Sports and travel	55122.8265
3	Electronic accessories	54337.5315
4	Fashion accessories	54305.8950
5	Home and lifestyle	53861.9130
6	Health and beauty	49193.7390

7. What is the city with the largest revenue?

```
select city, sum(total) Total_revenue from WalmartSalesData
group by city
order by Total_revenue desc
```

Output:

	city	Total_revenue
1	Naypyitaw	110568.7065
2	Yangon	106200.3705
3	Mandalay	106197.6720

8. What product line had the largest VAT?

```
select Product_line, AVG(Tax_5) Avg_VAT from WalmartSalesData
group by Product_line
order by Avg_VAT desc
```

Output:

	Product_line	Avg_VAT
1	Home and lifestyle	16.030331
2	Sports and travel	15.812629
3	Health and beauty	15.411572
4	Food and beverages	15.365310
5	Electronic accessories	15.220597
6	Fashion accessories	14.528061

9. Which branch sold more products than average product sold?

```
select Branch, sum(Quantity) Products_sold from WalmartSalesData
group by Branch
having (sum(Quantity) > (select avg(Quantity) from WalmartSalesData))
```

Output:

	Branch	Products_sold
1	A	1859
2	C	1831
3	B	1820

10. What is the most common product line by gender?

```
select gender, Product_line, count(gender) total_cnt from WalmartSalesData
group by gender, Product_line
order by total_cnt desc
```

Output:

	gender	Product_line	total_cnt
1	Female	Fashion accessories	96
2	Female	Food and beverages	90
3	Male	Health and beauty	88
4	Female	Sports and travel	88
5	Male	Electronic accessories	86
6	Female	Electronic accessories	84
7	Male	Food and beverages	84
8	Male	Fashion accessories	82
9	Male	Home and lifestyle	81
10	Female	Home and lifestyle	79
11	Male	Sports and travel	78
12	Female	Health and beauty	64

11. What is the average rating of each product line?

```
select Product_line, AVG(rating) Avg_Rating from WalmartSalesData
group by Product_line
```

Output:

	Product_line	Avg_Rating
1	Fashion accessories	7.029213
2	Home and lifestyle	6.837500
3	Electronic accessories	6.924705
4	Health and beauty	7.003289
5	Food and beverages	7.113218
6	Sports and travel	6.916265

- **Data analysis of sales-related queries**

1. How many unique customer types does the data have?

```
select distinct(customer_type), count(*) customer_type_cnt from WalmartSalesData
group by customer_type
```

Output:

	customer_type	customer_type_cnt
1	Member	501
2	Normal	499

2. How many unique payment methods does the data have?

```
select distinct(Payment), count(*) from WalmartSalesData
group by payment
```

Output:

	Payment	Payment_cnt
1	Credit card	311
2	Cash	344
3	Ewallet	345

3. What is the most common customer type?

```
select distinct(customer_type), count(*) customer_type_cnt from WalmartSalesData
group by customer_type
```

Output:

	customer_type	customer_type_cnt
1	Member	501
2	Normal	499

4. What is the gender of most of the customers?

```
select Gender, count(*) gender_cnt from WalmartSalesData
group by Gender
```

Output:

	Gender	gender_cnt
1	Male	499
2	Female	501

5. What is the gender distribution per branch?

```
select branch, gender, count(*) cnt_branchwise from WalmartSalesData
group by branch, gender
order by cnt_branchwise desc
```

Output:

	branch	gender	cnt_branchwise
1	A	Male	179
2	C	Female	178
3	B	Male	170
4	B	Female	162
5	A	Female	161
6	C	Male	150

6. Which day of the week has the best avg ratings?

```
select day_name, avg(Rating) Avg_Rating,count(*) No_of_rating_received from WalmartSalesData
group by day_name
order by Avg_Rating desc
```

Output:

	day_name	Avg_Rating	No_of_rating_received
1	Monday	7.153600	125
2	Friday	7.076258	139
3	Sunday	7.011278	133
4	Tuesday	7.003164	158
5	Saturday	6.901829	164
6	Thursday	6.889855	138
7	Wednesday	6.805594	143

- # Insights

1. **Payment Preferences:** Ewallet and Cash are the most favored payment methods, with credit cards as a close third.
2. **Product Sales Performance:** Fashion accessories lead in sales volume among all product lines.
3. **Revenue Trend:** A declining revenue trend has been observed since January, with March reflecting the lowest revenue.
4. **Sales Contribution by Product Lines:** Food & Beverages make the highest sales contribution, followed by Sports, Travel, and Electronic accessories.
5. **City-wise Sales Performance:** Naypyitaw demonstrates the highest sales revenue compared to other cities.
6. **Taxation Variance:** Home & Lifestyle products face the highest taxes, while fashion accessories incur the least.
7. **Gender-based Product Preferences:** Fashion accessories are the top choice among females, while health and beauty lead among males.
8. **Average Product Ratings:** All product lines boast consistently high average ratings, hovering around 7. Food and beverages receive the highest average rating.
9. **Day-wise Rating Patterns:** Average ratings peak on Mondays, followed by Friday and Sunday. Saturdays receive the highest number of ratings.
10. **Branch-wise Shopper Demographics:** Branch A attracts the highest male shoppers, whereas Branch C sees the highest number of female shoppers.

- # Recommendations

1. **Payment Method Optimization:** Encourage the promotion and ease of use of Ewallets and Credit cards as this will help reducing efforts for cash handling. Consider loyalty programs or discounts for customers using these payment methods.
2. **Product Line Enhancement:** Invest in marketing and product development for Fashion accessories, capitalizing on its current popularity to further boost sales.
3. **Revenue Recovery Strategies:** Investigate the factors contributing to the declining revenue trend since January and implement targeted marketing or promotional campaigns to reverse this trend.
4. **Strategic Product Placement:** Focus on promoting and strategically placing Food & Beverages, Sports, Travel, and Electronic accessories, as they contribute significantly to sales revenue.
5. **City-wise Marketing Strategies:** Tailor marketing strategies for Naypyitaw to leverage its high sales performance, potentially expanding market reach in other cities.
6. **Taxation Management:** Explore opportunities to optimize taxes on Home & Lifestyle products, and consider promotional strategies for fashion accessories to further stimulate sales.
7. **Gender-targeted Marketing:** Customize marketing campaigns based on gender preferences, promoting Fashion accessories for females and Health and Beauty for males.
8. **Product Rating Acknowledgment:** Acknowledge and promote the high average ratings, especially for Food and Beverages, to build trust and attract more customers.
9. **Strategic Day-wise Promotions:** Capitalize on the high average ratings on Mondays, Fridays, and Sundays by launching targeted promotions on these days. Consider special Saturday promotions to maximize ratings.
10. **Branch-specific Marketing:** Tailor marketing efforts at Branch A towards male shoppers and at Branch C towards female shoppers, creating a personalized shopping experience. Consider loyalty programs to enhance customer retention.

Appendix

Github link: https://github.com/Harikrishnan-Nair/SQL-Walmart-sales-data-analysis-project.git