VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY

UNIVERSITY OF INFORMATION TECHNOLOGY

FACULTY OF INFORMATION SYSTEMS

******

Final Project Report

TELCO CUSTOMER CHURN ETL, OLAP, AND DATA MINING SOLUTION

Lecturer: MSc. Nguyen Thi Kim Phung

Course: Decision Support & Business Intelligence

Class: MSIS4263.P21.CTTT

Student: Ngo Manh Khuong – 21522241

Ngo Cong Huan - 21520881

Ho Chi Minh city, June 2025

TABLE OF CONTENTS

I. SSIS process

Kaggle dataset: https://www.kaggle.com/code/anubhavgoyal10/customer-churn-prediction-eda-ann/input

Table 1: Dataset information

Column Name	Description
customerID	Unique identifier for each customer
gender	Customer's gender (Male, Female)
SeniorCitizen	Indicates if the customer is a senior (1 = Yes, 0 = No)
Partner	Whether the customer has a partner (Yes/No)
Dependents	Whether the customer has dependents (Yes/No)
tenure	Number of months the customer has stayed
PhoneService	Whether the customer has phone service (Yes/No)
MultipleLines	Whether the customer has multiple lines (Yes, No, No phone service)
InternetService	Type of internet (DSL, Fiber optic, No)
OnlineSecurity	Internet security service (Yes/No)
OnlineBackup	Online backup service (Yes/No)
DeviceProtection	Device protection plan (Yes/No)
TechSupport	Technical support plan (Yes/No)
StreamingTV	Streaming TV service (Yes/No)
StreamingMovies	Streaming movie service (Yes/No)
Contract	Contract type (Month-to-month, One year, Two year)
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)
PaymentMethod	Method of payment (e.g., Mailed check, Electronic check, Bank transfer)
MonthlyCharges	Current monthly charges
TotalCharges	Total amount charged to the customer
Churn	Target variable: whether the customer has churned (Yes/No)

Descriptions of Dimensions:

+ Dim_Customer

Column	Data Type	Description
customerID	nvarchar(255)	Unique identifier for each customer (acts as primary key).
gender	nvarchar(255)	Gender of the customer (Male, Female).
SeniorCitizen	float	Indicates if the customer is a senior citizen (e.g., 0 = No, 1 = Yes).
Partner	nvarchar(255)	Whether the customer has a partner (Yes/No).
Dependents	nvarchar(255)	Whether the customer has dependents (Yes/No).
PaperlessBilling	nvarchar(255)	Whether the customer uses paperless billing (Yes/No).

Purpose: Stores demographic and billing behavior details of customers.
BI Contribution:

- Enables segmentation of churn by gender, senior status, partnership status, and billing preferences.

- Answers questions like:

  o *Are senior citizens more likely to churn?*

  o *Do customers with dependents stay longer?*

  o *Is paperless billing associated with lower churn?*

+ Dim_InternetService

Column	Data Type	Description
InternetServiceID	int	Surrogate key for the internet service dimension.
InternetService	nvarchar(255)	Type of internet service (DSL, Fiber optic, No...).
InternetAvailability	nvarchar(255)	General category: Has Internet, No Internet.

Purpose: Describes the type and availability of internet service used.
BI Contribution:

- Help identify churn patterns based on service availability.

- Distinguishes between users with No Internet vs. DSL or Fiber.

- Answers questions like:

    o *Does fiber optic service lead to better customer retention?*

    o *Are customers without internet more likely to leave?*

+ Dim_PaymentMethod

Column	Data Type	Description
PaymentMethodID	int	Surrogate key for the payment method.
PaymentCategory	nvarchar(255)	Category of payment method: Automatic or Manual.
PaymentMethod	nvarchar(255)	Specific payment method (Electronic check, Bank transfer, etc.).

Purpose: Stores the specific and grouped methods of customer payments.
BI Contribution:

- Enables analysis of how payment habits affect churn.

- Splits into Automatic vs. Manual, helping understand convenience vs. churn.

- Answers questions like:

    o *Do automatic payments reduce churn rates?*

    o *Which payment methods are common among loyal customers?*

+ Dim_AdditionalServices

Column	Data Type	Description
AdditionalServicesID	int	Surrogate key.
PhoneService	nvarchar(255)	Whether the customer has phone service.
MultipleLines	nvarchar(255)	Multiple phone lines (Yes/No).
OnlineSecurity	nvarchar(255)	Whether online security is enabled.
OnlineBackup	nvarchar(255)	Whether online backup is enabled.
DeviceProtection	nvarchar(255)	Whether device protection is enabled.

Column	Data Type	Description
TechSupport	nvarchar(255)	Whether tech support is included.
StreamingTV	nvarchar(255)	Whether streaming TV is enabled.
StreamingMovies	nvarchar(255)	Whether streaming movies are enabled.

Purpose: Captures whether customers subscribed to value-added services (phone, streaming, backup, etc.).
BI Contribution:

- Helps identify product bundles that retain customers longer.

- Correlates specific services (e.g., streamingTV) with loyalty.

- Answers questions like:

    o *Does having tech support or online security reduce churn?*

    o *Which service combinations have the highest retention?*

+ Dim_Contract

Column	Data Type	Description
contractID	int	Surrogate key.
ContractGroup	nvarchar(255)	Grouping (Short-term, Long-term).
Contract	nvarchar(255)	Specific contract type (Month-to-month, One year, Two year).

Purpose: Contains contract types and their length classification.
BI Contribution:

- Shows how contract commitment level affects churn.

- Groups customers by short-term vs. long-term behavior.

- Answers questions like:

    o *Do month-to-month customers churn more than yearly ones?*

    o *Is long-term commitment effective in reducing churn?*

+ Dim_Tenure

Column	Data Type	Description
tenureID	int	Surrogate key.
tenureRange	nvarchar(255)	Bucketed tenure range (e.g., 0–12 months, 13–24 months, etc.).
tenureGroup	nvarchar(255)	Grouping of tenure: Short-term, Medium-term, Long-term.
tenure	int	Actual number of months the customer has stayed.

Purpose: Captures how long a customer has been with the company, both as raw months and grouped buckets.

BI Contribution:

- Crucial for understanding customer lifecycle.

- Helps identify tenure-based retention strategies.

- Answers questions like:

    o *Are newer customers more likely to churn?*

    o *Which tenure group is most loyal or most at risk?*

+ Dim_Churn

Column	Data Type	Description
ChurnID	int	Surrogate key.
Churn	nvarchar(255)	Indicates whether the customer churned (Yes, No).

Purpose: Indicates whether a customer has churned.

BI Contribution:

- Used as a label or target metric in churn reporting, KPIs, and predictive analytics.

- Enables side-by-side comparisons of churned vs. retained groups.

- Answers questions like:

    o *What percentage of customers churned last month?*

    o *What are the profiles of customers who tend to churn?*

+ CustomerChurnFact (Fact Table)

Column	Data Type	Description
FactID	int (PK)	Surrogate primary key for the fact table.
customerID	nvarchar(255)	Foreign key to Dim_Customer.
InternetServiceID	int	Foreign key to Dim_InternetService.
PaymentMethodID	int	Foreign key to Dim_PaymentMethod.
AdditionalServicesID	int	Foreign key to Dim_AdditionalServices.
contractID	int	Foreign key to Dim_Contract.
tenureID	int	Foreign key to Dim_Tenure.
ChurnID	int	Foreign key to Dim_Churn.
MonthlyCharges	float	The amount charged to the customer monthly.
TotalCharges	float	The total amount charged to the customer.

Describe the SSIS process
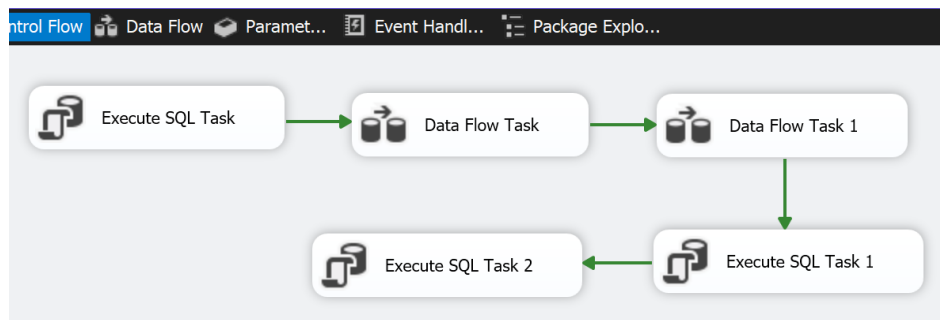
Figure 1: Control Flow structure

In the first SQL task execution, this task deletes all data from the tables and reset the IDs for selected table.
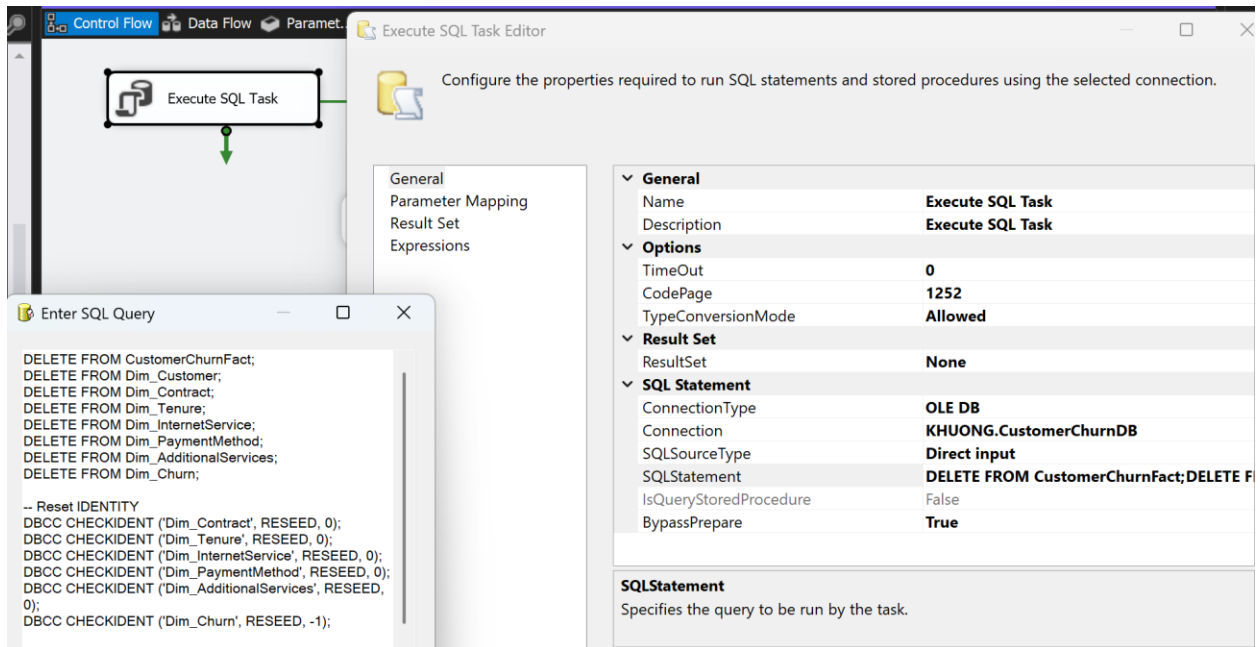
Figure 2: First SQL task execution

In the second SQL task execution, we add constraints for all tables.
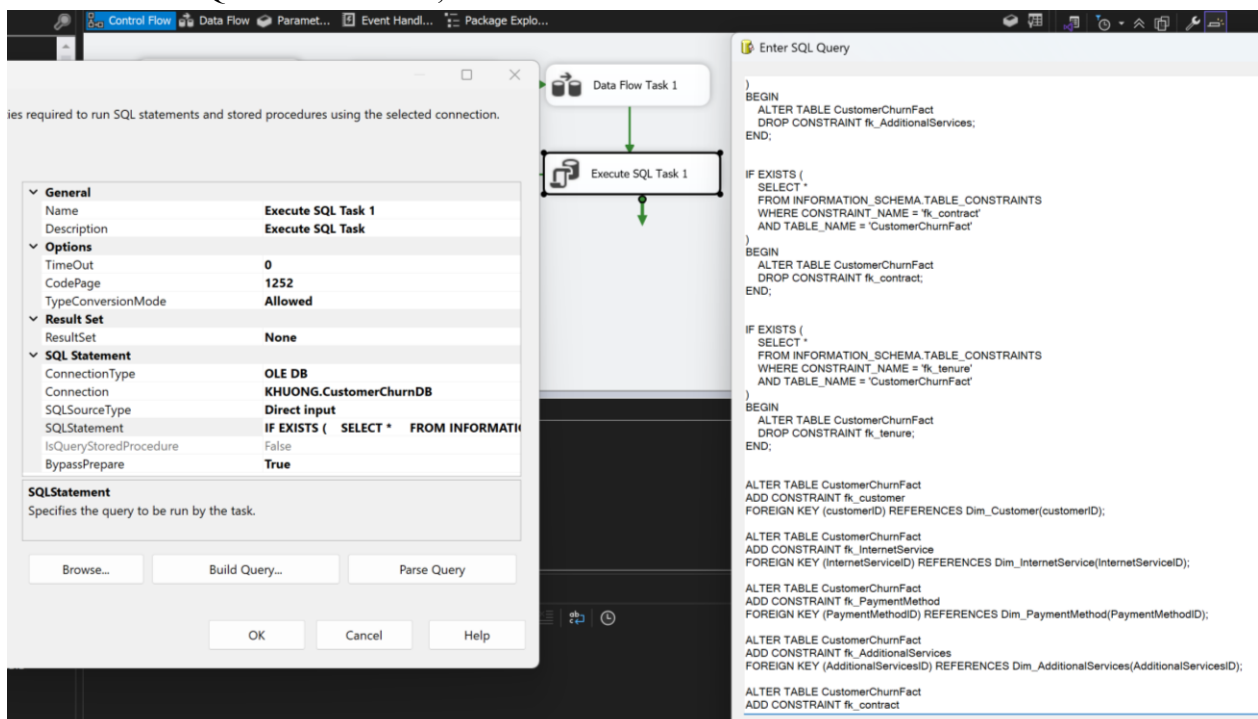
Figure 3: Second SQL task execution

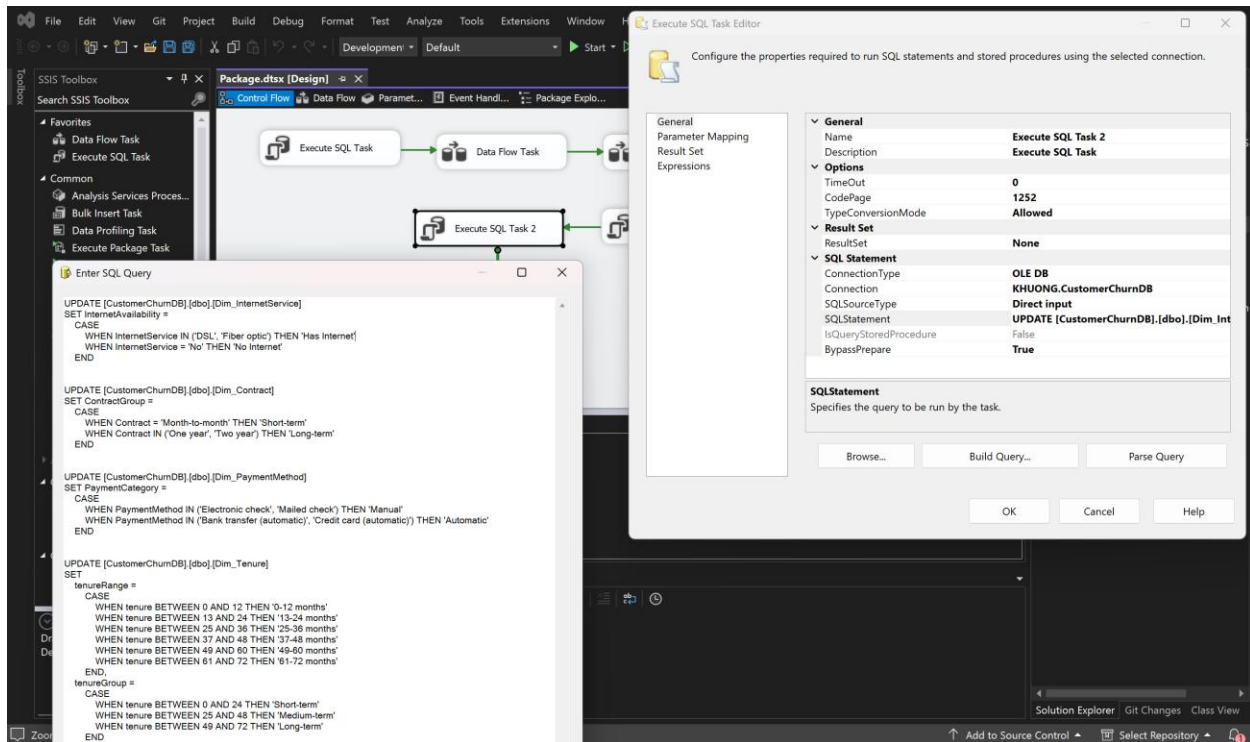In the last SQL task execution, we grouped the data for each table.
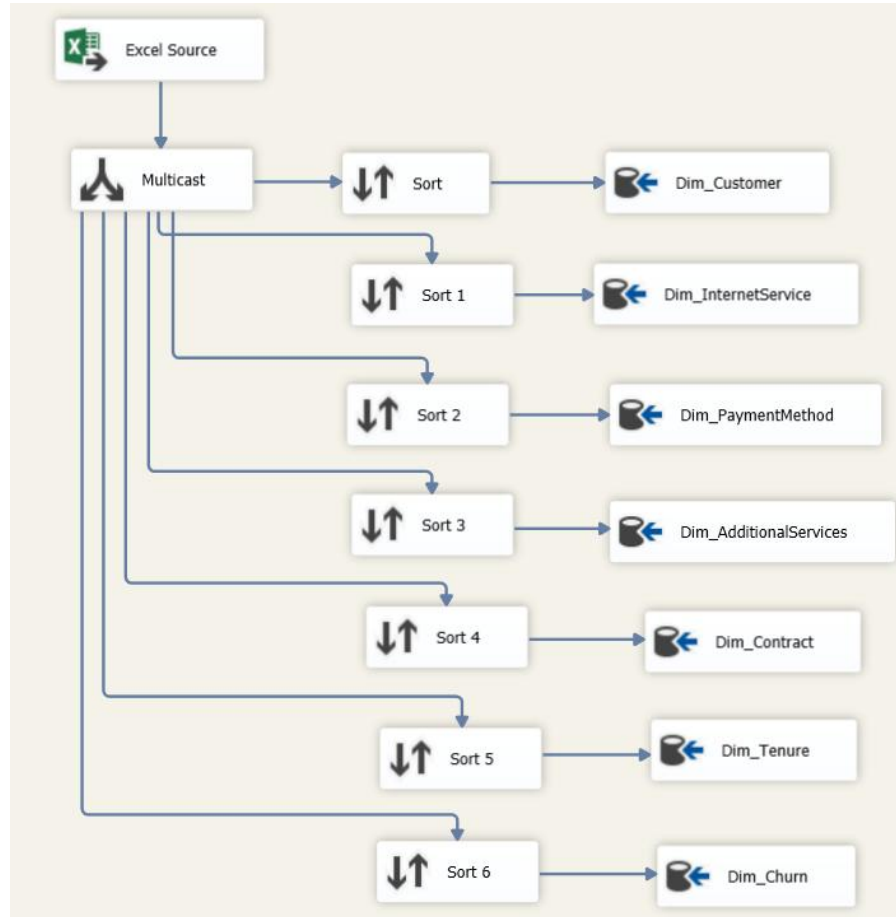

Figure 4: Last SQL task execution
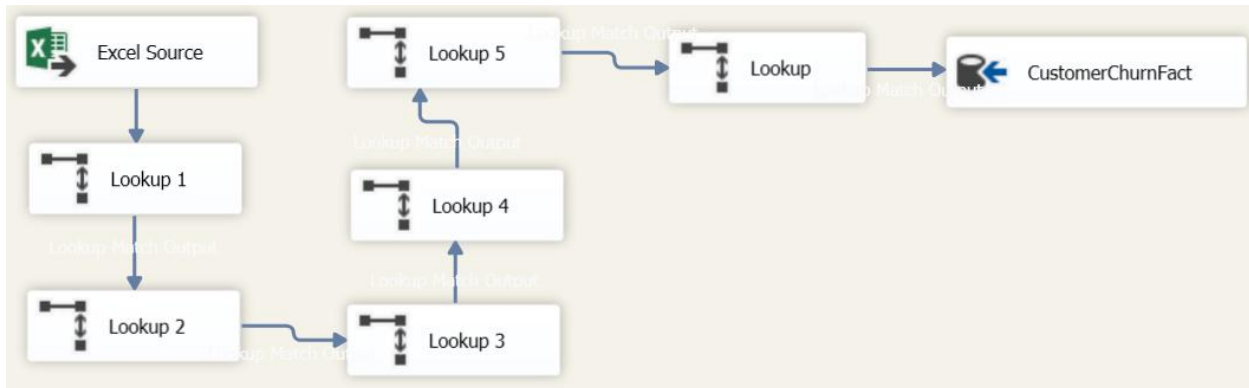
Figure 5: First Data Flow Task

Figure 6: Second Data Flow Task

II. Analyzing and reporting processes

a) Describe the schema (star/snow)

Schema Type: Star Schema

- There is one central fact table: CustomerChurnFact.

- Surrounding it are several dimension tables:

    o Dim_Customer

    o Dim_InternetService

    o Dim_PaymentMethod

    o Dim_AdditionalServices

    o Dim_Contract

    o Dim_Tenure

    o Dim_Churn

- All dimension tables are directly linked to the fact table via foreign keys (no hierarchies or snowflaked sub-dimensions).

- There is no further normalization of dimensions into sub-dimensions (which would be typical of a snowflake schema).
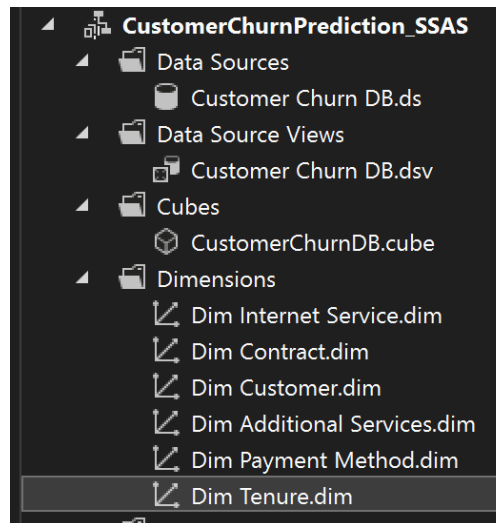
b) Process of building a cube
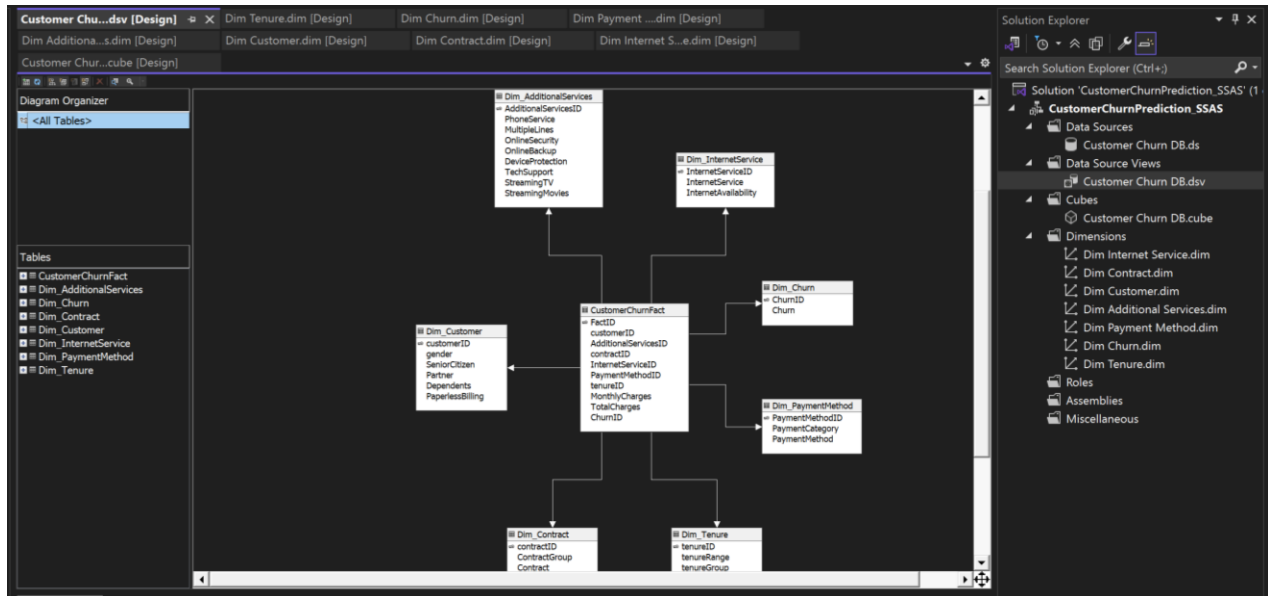

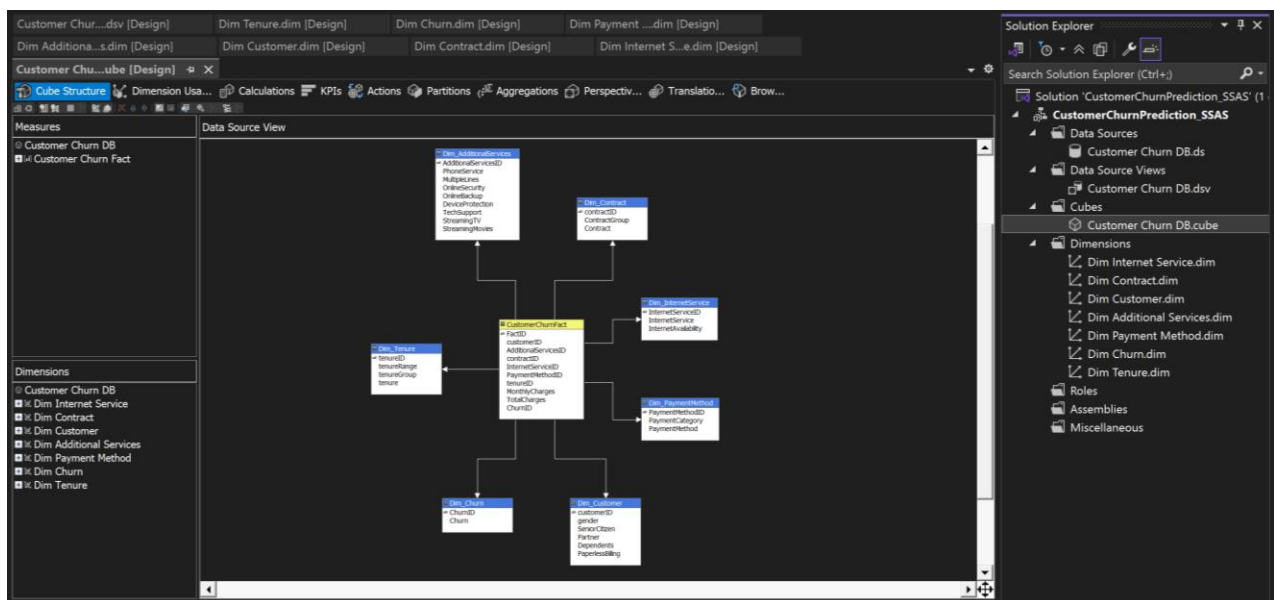*Figure 7: Structure of SSAS*

Figure 8: Data Source View

Figure 9: Cube

c) Using analysis service (manual), creating reports (Power BI), analysis service (MDX) and Pivot table (Excel)

- Below are 10 Business intelligence scenarios:

1. TenureGroup Customer Counts

Row (TenureGroup), Column (Churn), Values (Count of CustomerID)

The retention team is trying to identify new subscribers (Short-term tenure) vs. long-term loyalists. By examining churn by tenure group, they can launch loyalty campaigns to retain those most likely to churn early.
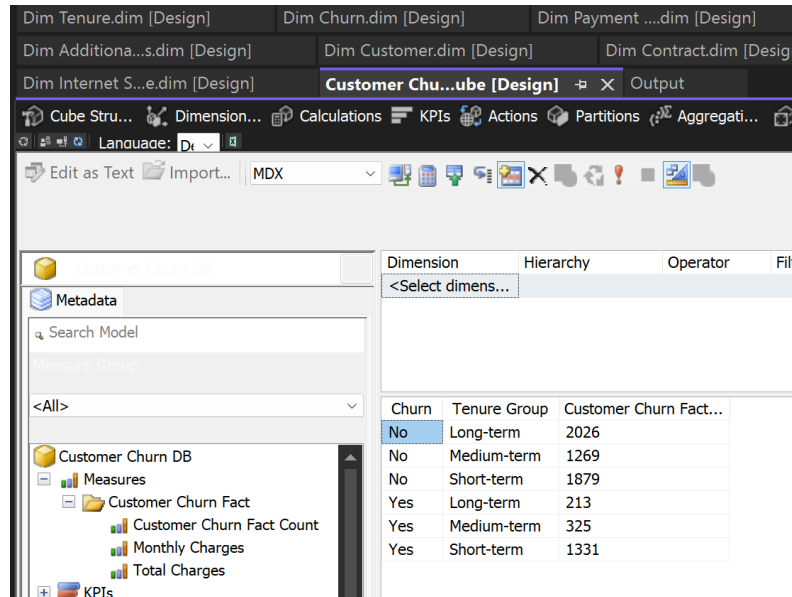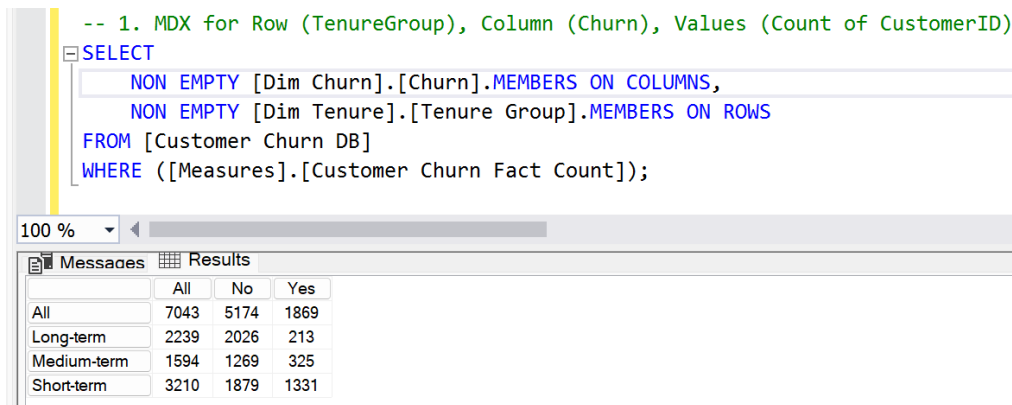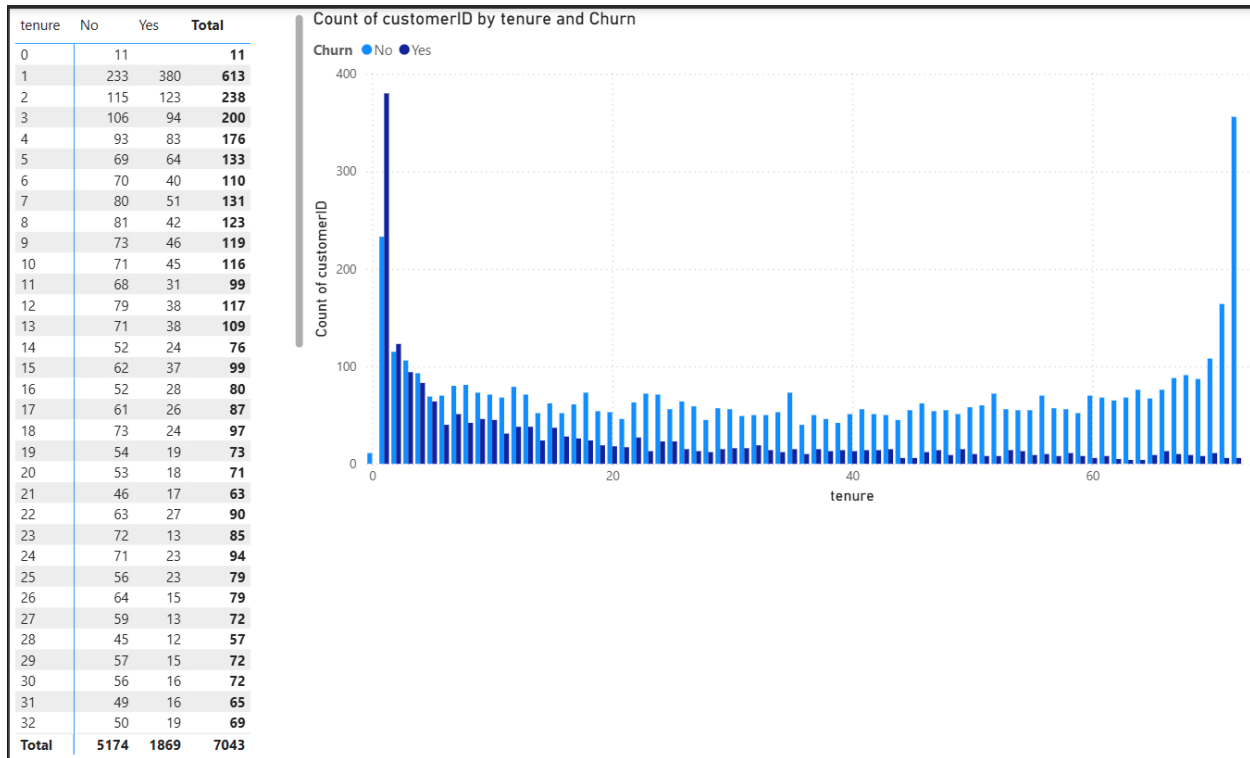

Figure 10: SSAS

```
-- 1. MDX for Row (TenureGroup), Column (Churn), Values (Count of CustomerID)
SELECT
    NON EMPTY [Dim Churn].[Churn].MEMBERS ON COLUMNS,
    NON EMPTY [Dim Tenure].[Tenure Group].MEMBERS ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Customer Churn Fact Count]);
```

	All	No	Yes
All	7043	5174	1869
Long-term	2239	2026	213
Medium-term	1594	1269	325
Short-term	3210	1879	1331

Figure 11: MDX

tenure	No	Yes	Total
0	11		11
1	233	380	613
2	115	123	238
3	106	94	200
4	93	83	176
5	69	64	133
6	70	40	110
7	80	51	131
8	81	42	123
9	73	46	119
10	71	45	116
11	68	31	99
12	79	38	117
13	71	38	109
14	52	24	76
15	62	37	99
16	52	28	80
17	61	26	87
18	73	24	97
19	54	19	73
20	53	18	71
21	46	17	63
22	63	27	90
23	72	13	85
24	71	23	94
25	56	23	79
26	64	15	79
27	59	13	72
28	45	12	57
29	57	15	72
30	56	16	72
31	49	16	65
32	50	19	69
Total	5174	1869	7043

Figure 12: PowerBI

Count of customerID	Column L		
Row Labels	No	Yes	Grand Total
0	11		11
1	233	380	613
2	115	123	238
3	106	94	200
4	93	83	176
5	69	64	133
6	70	40	110
7	80	51	131
8	81	42	123
9	73	46	119
10	71	45	116
11	68	31	99
12	79	38	117
13	71	38	109
14	52	24	76
15	62	37	99
16	52	28	80
17	61	26	87
18	73	24	97
19	54	19	73
20	53	18	71
21	46	17	63

Figure 13: Excel

2. InternetService and ContractGroup Monthly Charges

Row (InternetService), Column (ContractGroup), Values (Average of MonthlyCharges)

The pricing team is reviewing high-cost internet services and contract types (e.g., Long-term fiber internet users) to consider discounts or promotional offers next quarter.

Figure 14: SSAS

```
--2. MDX for Row (InternetService), Column (ContractGroup), Values (Average of
WITH
MEMBER [Measures].[Avg Monthly Charges] AS
    IIF(
        COUNT(
            NONEMPTY(
                [Dim Customer].[Customer ID].[Customer ID].MEMBERS,
                ([Measures].[Monthly Charges],
                 [Dim Internet Service].[Internet Service].CurrentMember,
                 [Dim Contract].[Contract].CurrentMember)
            )
        ) = 0,
        NULL,
        SUM(
            NONEMPTY(
                [Dim Customer].[Customer ID].[Customer ID].MEMBERS,
                ([Measures].[Monthly Charges],
```

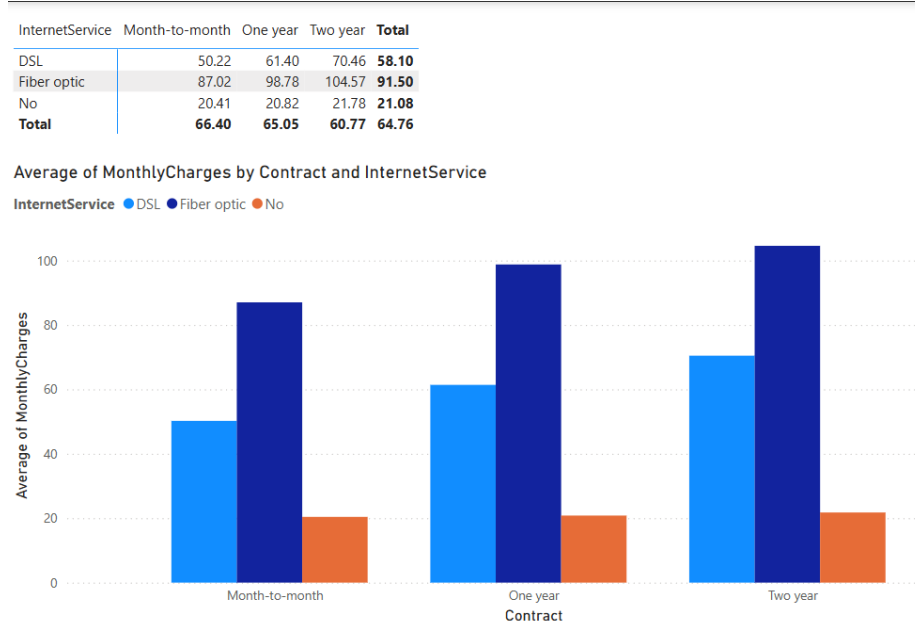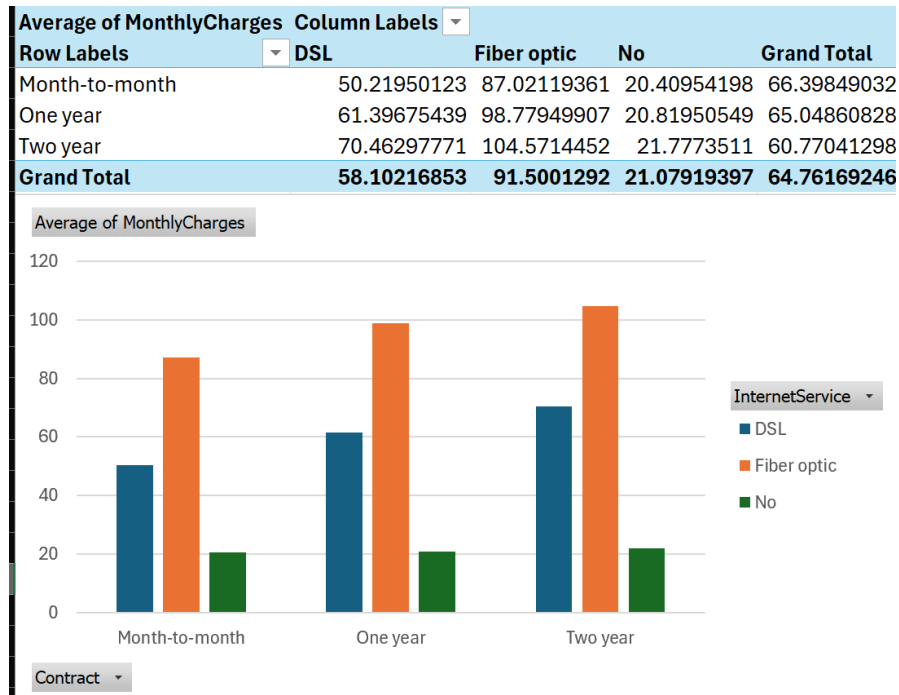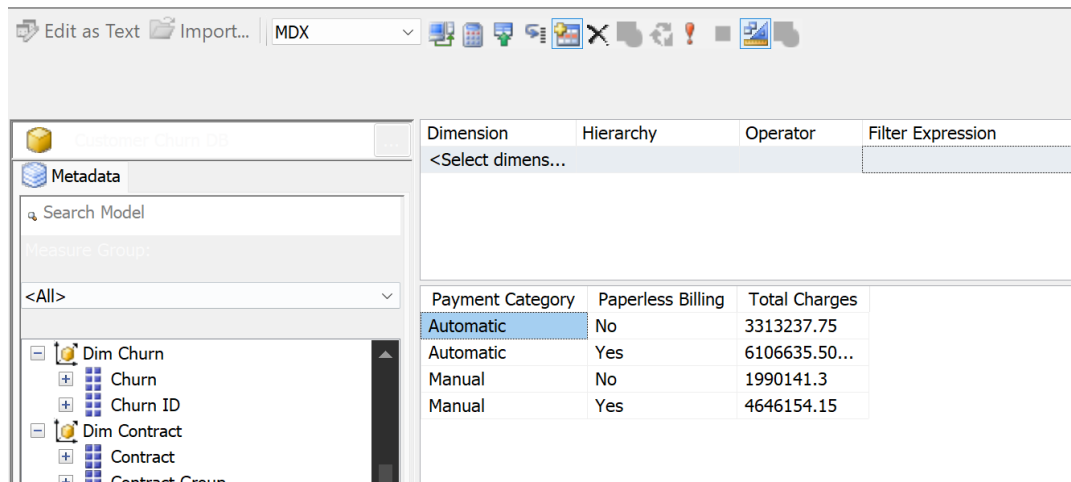|	DSL	Fiber optic	No	All
Month-to-month	50.2195012264923	87.0211936090227	20.4095419847328	66.3984903225806
One year	61.396754385965	98.7794990723563	20.8195054945055	65.0486082824168
Two year	70.4629777070063	104.571445221445	21.7773510971787	60.7704129793511
All	58.1021685254027	91.5001291989665	21.0791939711664	64.7616924605992

Figure 15: MDX

InternetService	Month-to-month	One year	Two year	Total
DSL	50.22	61.40	70.46	**58.10**
Fiber optic	87.02	98.78	104.57	**91.50**
No	20.41	20.82	21.78	**21.08**
Total	**66.40**	**65.05**	**60.77**	**64.76**

Average of MonthlyCharges by Contract and InternetService

InternetService ● DSL ● Fiber optic ● No

Figure 16: PowerBI

Average of MonthlyCharges	Column Labels			
Row Labels	DSL	Fiber optic	No	Grand Total
Month-to-month	50.21950123	87.02119361	20.40954198	66.39849032
One year	61.39675439	98.77949907	20.81950549	65.04860828
Two year	70.46297771	104.5714452	21.7773511	60.77041298
Grand Total	58.10216853	91.5001292	21.07919397	64.76169246

Figure 17: Excel

3. PaymentCategory and PaperlessBilling Total Charges
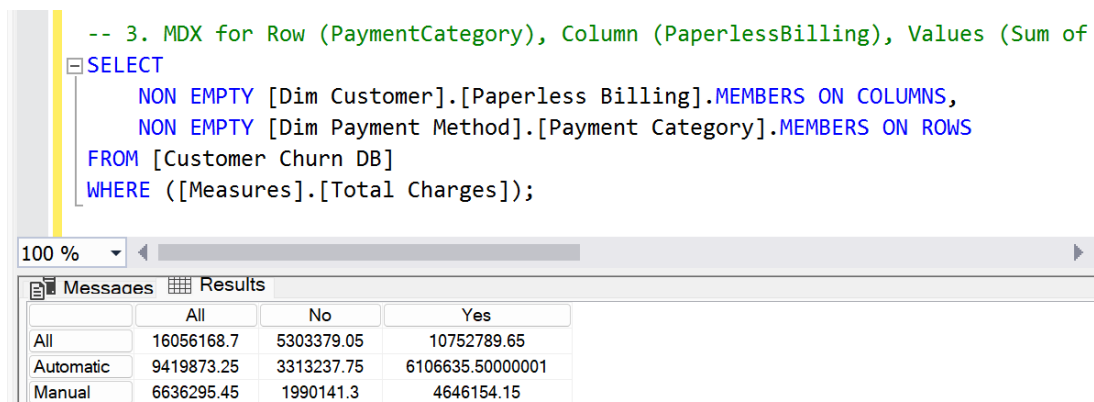
Row (PaymentCategory), Column (PaperlessBilling), Values (Sum of TotalCharges)

The finance team wants to forecast total revenue and finds that customers using automatic payments and paperless billing generate more revenue. They will promote paperless billing to boost revenue.

Figure 18: SSAS

```
-- 3. MDX for Row (PaymentCategory), Column (PaperlessBilling), Values (Sum of
SELECT
    NON EMPTY [Dim Customer].[Paperless Billing].MEMBERS ON COLUMNS,
    NON EMPTY [Dim Payment Method].[Payment Category].MEMBERS ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Total Charges]);
```

	All	No	Yes
All	16056168.7	5303379.05	10752789.65
Automatic	9419873.25	3313237.75	6106635.50000001
Manual	6636295.45	1990141.3	4646154.15

Figure 19: MDX

PaymentMethod	No	Yes	Total
Bank transfer (automatic)	1,650,285.35	3,097,994.55	**4,748,279.90**
Credit card (automatic)	1,662,952.40	3,008,640.95	**4,671,593.35**
Electronic check	1,112,114.10	3,832,789.15	**4,944,903.25**
Mailed check	878,027.20	813,365.00	**1,691,392.20**
Total	**5,303,379.05**	**10,752,789.65**	**16,056,168.70**

Figure 20: PowerBI

Sum of TotalCharges	Column Labels		
Row Labels	No	Yes	Grand Total
Bank transfer (automatic)	1650285.35	3097994.55	4748279.9
Credit card (automatic)	1662952.4	3008640.95	4671593.4
Electronic check	1112114.1	3832789.15	4944903.3
Mailed check	878027.2	813365	1691392.2
Grand Total	**5303379.05**	**10752789.7**	**16056169**

Figure 21: Excel

4. Contract and InternetAvailability Customer Counts

Row (Contract), Column (InternetAvailability), Values (Count of CustomerID)
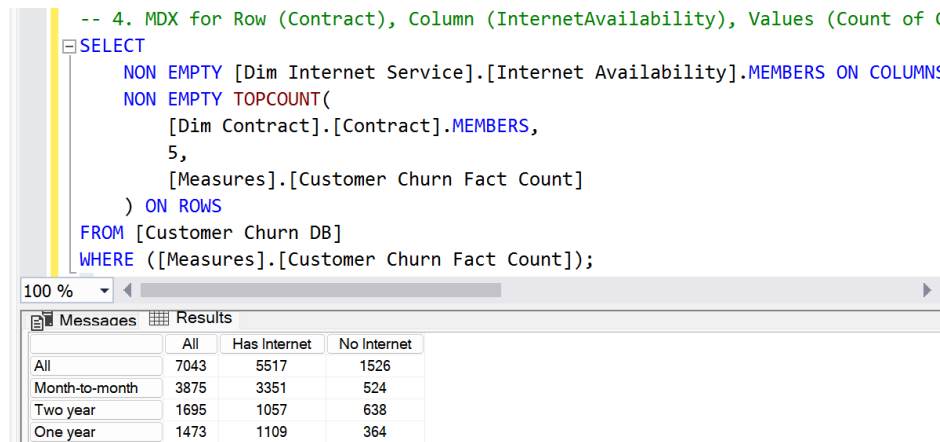
Customer service wants to know which contract types are popular among users with internet. This helps in targeting upgrades or bundles for contracts with frequent complaints.

Figure 22: SSAS

```
-- 4. MDX for Row (Contract), Column (InternetAvailability), Values (Count of (
SELECT
    NON EMPTY [Dim Internet Service].[Internet Availability].MEMBERS ON COLUMNS
    NON EMPTY TOPCOUNT(
        [Dim Contract].[Contract].MEMBERS,
        5,
        [Measures].[Customer Churn Fact Count]
    ) ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Customer Churn Fact Count]);
```

	All	Has Internet	No Internet
All	7043	5517	1526
Month-to-month	3875	3351	524
Two year	1695	1057	638
One year	1473	1109	364

Figure 23: MDX

Contract	DSL	Fiber optic	No	Total
Month-to-month	1223	2128	524	**3875**
One year	570	539	364	**1473**
Two year	628	429	638	**1695**
Total	**2421**	**3096**	**1526**	**7043**

Figure 24: PowerBI

Count of customerID	Column Labels			
Row Labels	**DSL**	**Fiber optic**	**No**	**Grand Total**
Month-to-month	1223	2128	524	3875
One year	570	539	364	1473
Two year	628	429	638	1695
Grand Total	**2421**	**3096**	**1526**	**7043**
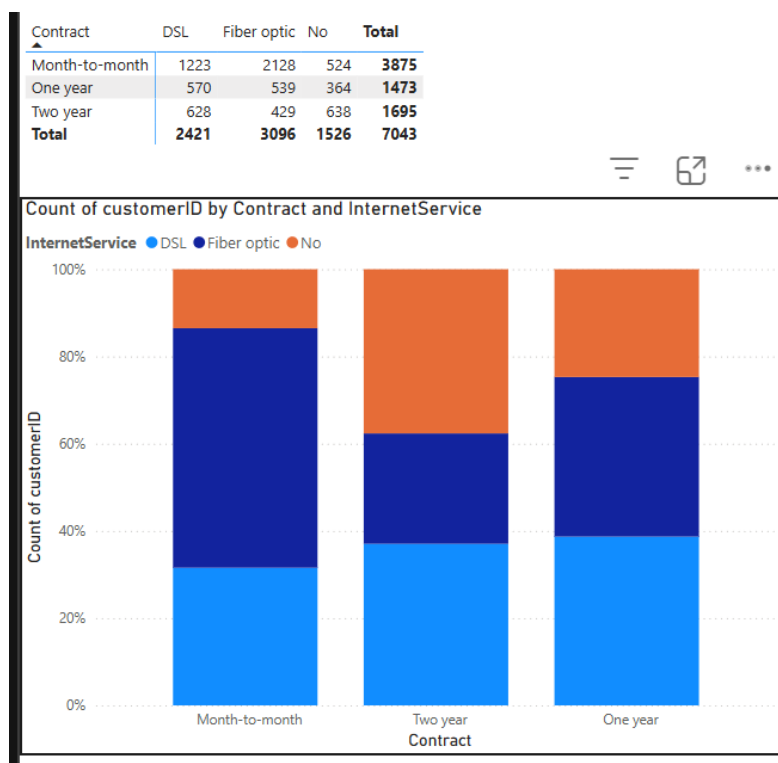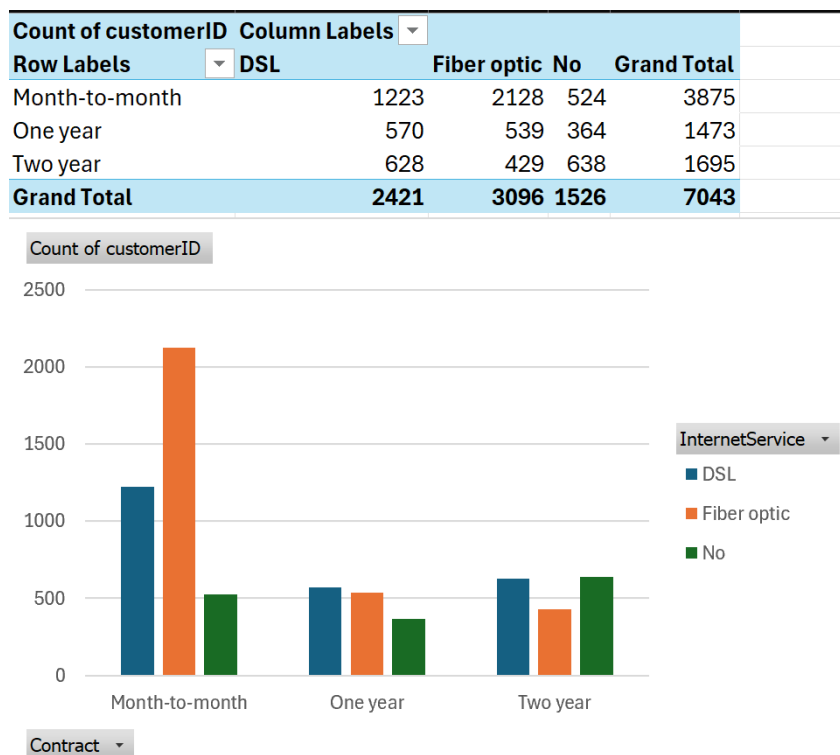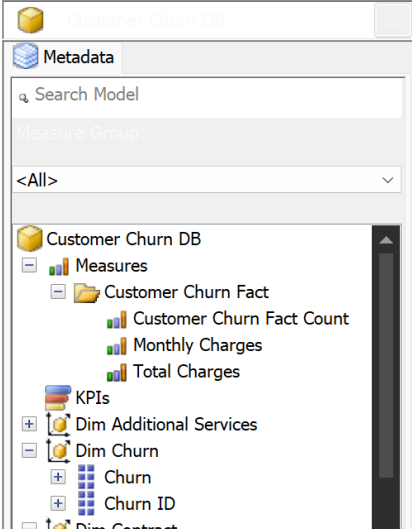
Figure 25: Excel

5. TenureRange and SeniorCitizen Total Charges

Row (TenureRange), Column (SeniorCitizen), Values (Average of TotalCharges)

Product development sees that senior citizens often have lower total charges and can now design affordable packages tailored for this demographic, especially newer customers.

Figure 26: SSAS

Figure 27: MDX

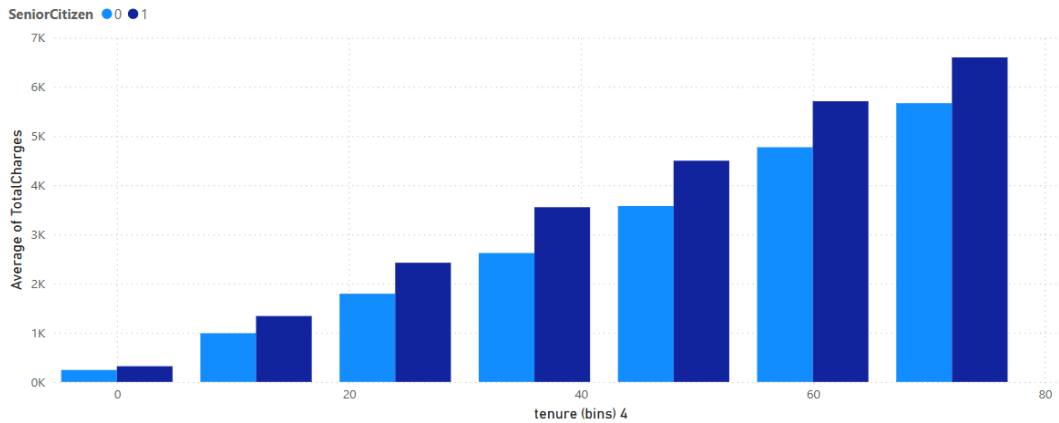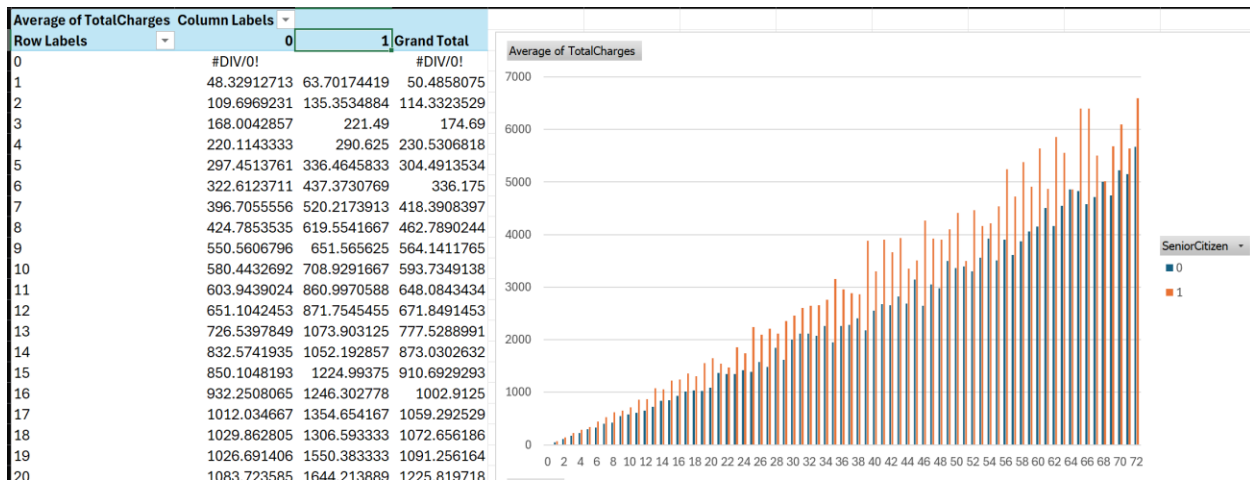| SeniorCitizen	0	12	24	36	48	60	72	**Total**
0	242.70	989.82	1,793.61	2,621.13	3,576.86	4,772.30	5,668.43	**2,181.09**
1	318.95	1,339.66	2,423.80	3,552.29	4,498.70	5,708.31	6,599.39	**2,810.47**
Total	254.15	1,044.28	1,913.75	2,774.24	3,733.13	4,926.78	5,812.45	**2,283.30**

Figure 28: PowerBI

Figure 29: Excel

6. InternetService and Dependents Monthly Charges

Row (InternetService), Column (Dependents), Values (Average of MonthlyCharges)

Marketing identifies single customers without dependents paying high monthly fees for internet. These customers may churn, so a discount plan can help retain them.

Figure 30: SSAS

```
--6. MDX for Row (InternetService), Column (Dependents),
-- Values (Average of MonthlyCharges)
WITH
MEMBER [Measures].[Avg Monthly Charges] AS
    IIF(
        COUNT(
            NONEMPTY(
                [Dim Customer].[Customer ID].[Customer ID].MEMBERS,
                ([Measures].[Monthly Charges],
                [Dim Customer].[Dependents].CurrentMember,
                [Dim Internet Service].[Internet Service].CurrentMember)
            )
        ) = 0,
```

100 %

Messages   Results

	No	Yes	All
DSL	56.4437809405941	61.4313043478261	58.1021685254027
Fiber optic	90.7188989317996	94.3725075528701	91.5001291989665
No	20.9534541336353	21.251866251944	21.0791939711664
All	67.0028076221365	59.5221563981042	64.7616924605992

Figure 31: MDX

InternetService	No	Yes	Total
DSL	56.44	61.43	**58.10**
Fiber optic	90.72	94.37	**91.50**
No	20.95	21.25	**21.08**
Total	**67.00**	**59.52**	**64.76**

Average of MonthlyCharges by InternetService and Dependents

Dependents ● No ● Yes

Figure 32: PowerBI

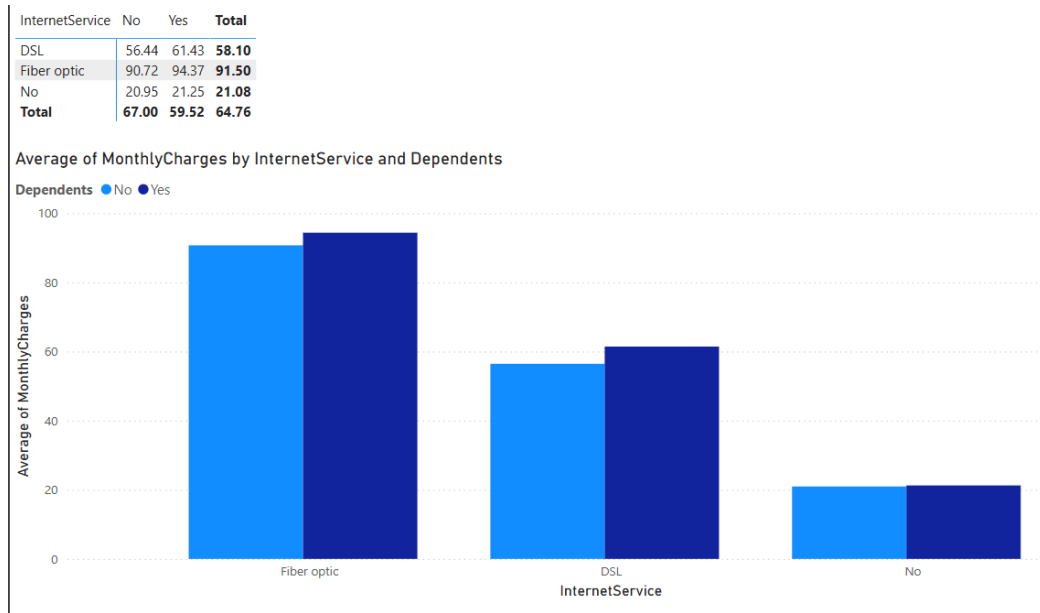| Average of MonthlyCharges	Column Labels		
Row Labels	**No**	**Yes**	**Grand Total**
DSL	56.44378094	61.43130435	58.10216853
Fiber optic	90.71889893	94.37250755	91.5001292
No	20.95345413	21.25186625	21.07919397
Grand Total	**67.00280762**	**59.5221564**	**64.76169246**

Average of MonthlyCharges

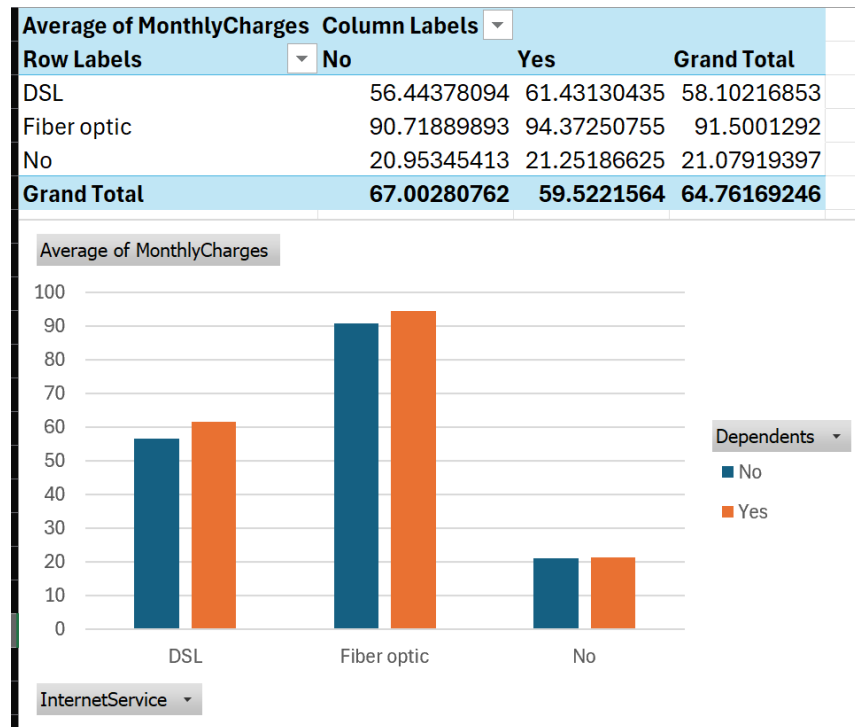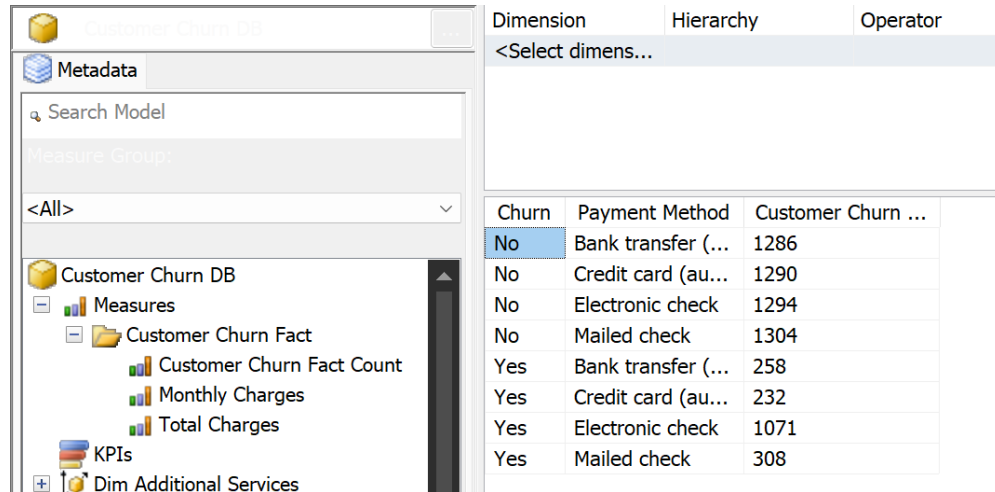Dependents
■ No
■ Yes

InternetService

Figure 33: Excel

7. PaymentMethod Customer Counts

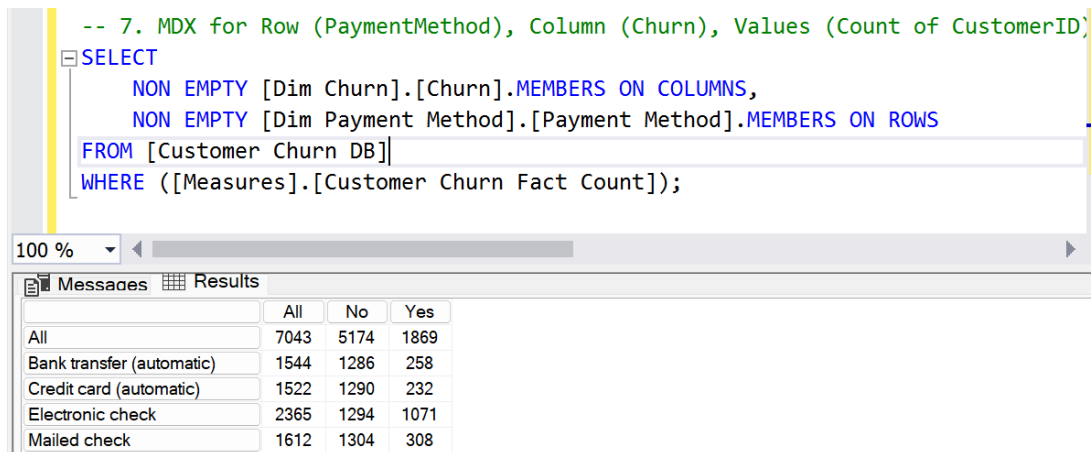**Row (PaymentMethod), Column (Churn), Values (Count of CustomerID)**

Operations want to find which payment methods are preferred or problematic. For example, if manual check users churn more, they can promote automatic payments via incentives.

Figure 34: SSAS

```
-- 7. MDX for Row (PaymentMethod), Column (Churn), Values (Count of CustomerID)
SELECT
    NON EMPTY [Dim Churn].[Churn].MEMBERS ON COLUMNS,
    NON EMPTY [Dim Payment Method].[Payment Method].MEMBERS ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Customer Churn Fact Count]);
```

	All	No	Yes
All	7043	5174	1869
Bank transfer (automatic)	1544	1286	258
Credit card (automatic)	1522	1290	232
Electronic check	2365	1294	1071
Mailed check	1612	1304	308

Figure 35: MDX

PaymentMethod	No	Yes	Total
Bank transfer (automatic)	1286	258	**1544**
Credit card (automatic)	1290	232	**1522**
Electronic check	1294	1071	**2365**
Mailed check	1304	308	**1612**
Total	**5174**	**1869**	**7043**

Count of customerID by PaymentMethod and Churn

Figure 36: PowerBI

Count of customerID	Column Labels		
Row Labels	No	Yes	Grand Total
Bank transfer (automatic)	1286	258	1544
Credit card (automatic)	1290	232	1522
Electronic check	1294	1071	2365
Mailed check	1304	308	1612
Grand Total	**5174**	**1869**	**7043**

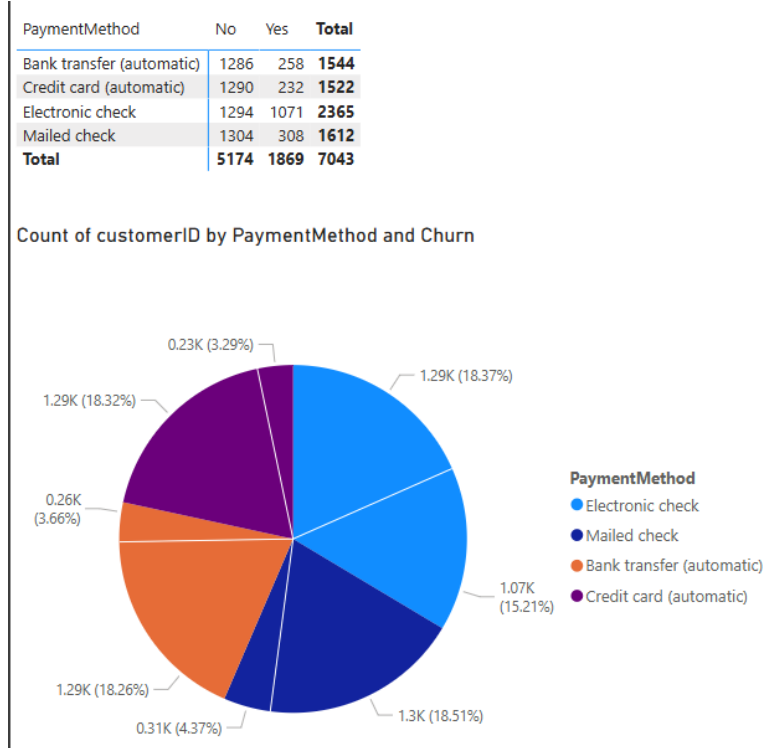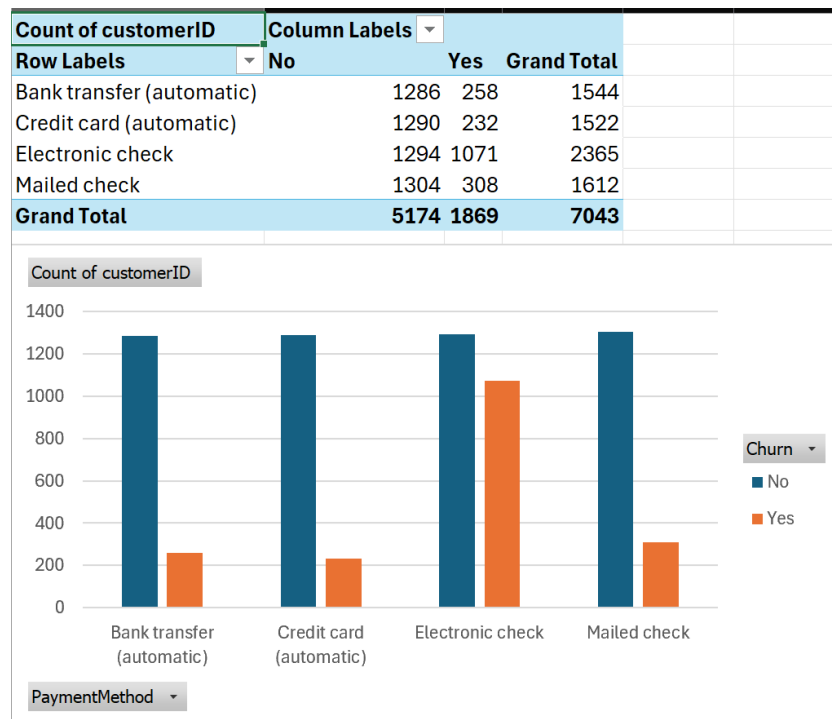*Figure 37: Excel*

8. ContractGroup and Gender Total Charges

Row (ContractGroup), Column (Gender), Values (Sum of TotalCharges)

Sales examines gender-based revenue patterns to target high-revenue customers (e.g., females in long-term contracts) with exclusive promotions or perks.

Figure 38: SSAS

```
-- 8. MDX for Row (ContractGroup), Column (Gender), Values (Sum of TotalCharges
SELECT
    NON EMPTY [Dim Customer].[Gender].MEMBERS ON COLUMNS,
    NON EMPTY [Dim Contract].[Contract Group].MEMBERS ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Total Charges]);
```

	All	Female	Male
All	16056168.7	7952354.20000001	8103814.49999999
Long-term	10750307.2	5346378.40000001	5403928.79999999
Short-term	5305861.5	2605975.8	2699885.7

Figure 39: MDX

Contract	Female	Male	Total
Month-to-month	2,605,975.80	2,699,885.70	**5,305,861.50**
One year	2,248,225.90	2,218,827.60	**4,467,053.50**
Two year	3,098,152.50	3,185,101.20	**6,283,253.70**
Total	**7,952,354.20**	**8,103,814.50**	**16,056,168.70**

Sum of TotalCharges by Contract and gender

gender ● Female ● Male

Figure 40: PowerBI

Sum of TotalCharges	Column Labels		
Row Labels	Female	Male	Grand Total
Month-to-month	2605975.8	2699885.7	5305861.5
One year	2248225.9	2218827.6	4467053.5
Two year	3098152.5	3185101.2	6283253.7
Grand Total	**7952354.2**	**8103814.5**	**16056168.7**

Sum of TotalCharges

gender
■ Female
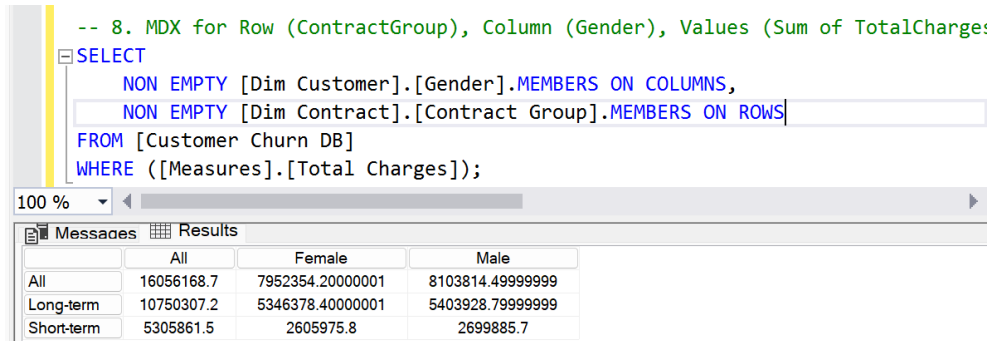■ Male

Contract

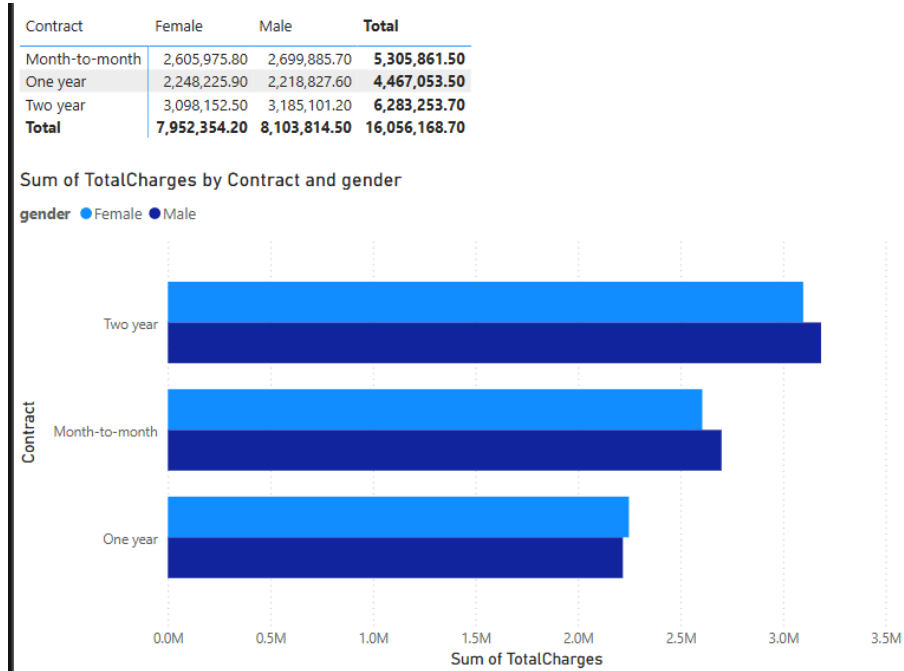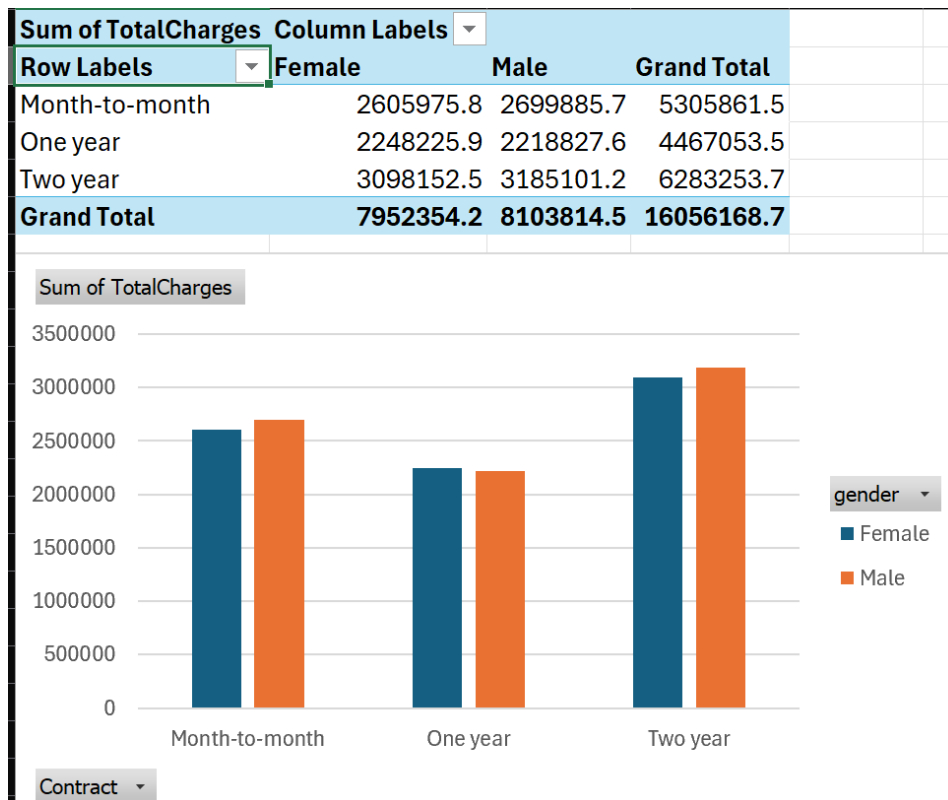*Figure 41: Excel*

9. Tenure and Partner Monthly Charges

Row (Tenure), Column (Partner), Values (Average of MonthlyCharges)

The analytics team explores if customers with partners and long tenure pay more. If so, they can create a family loyalty program that rewards staying longer as a couple.
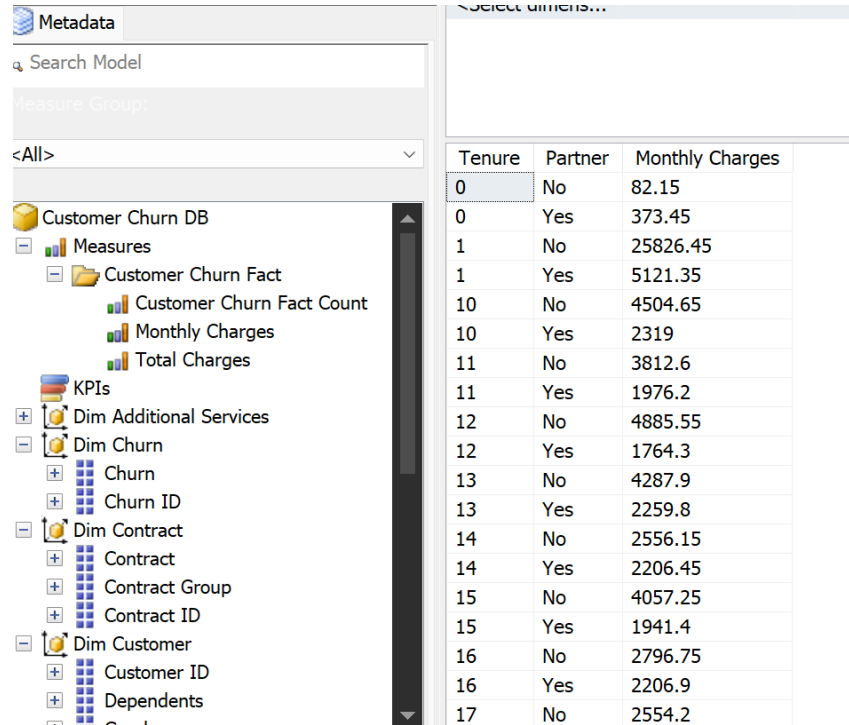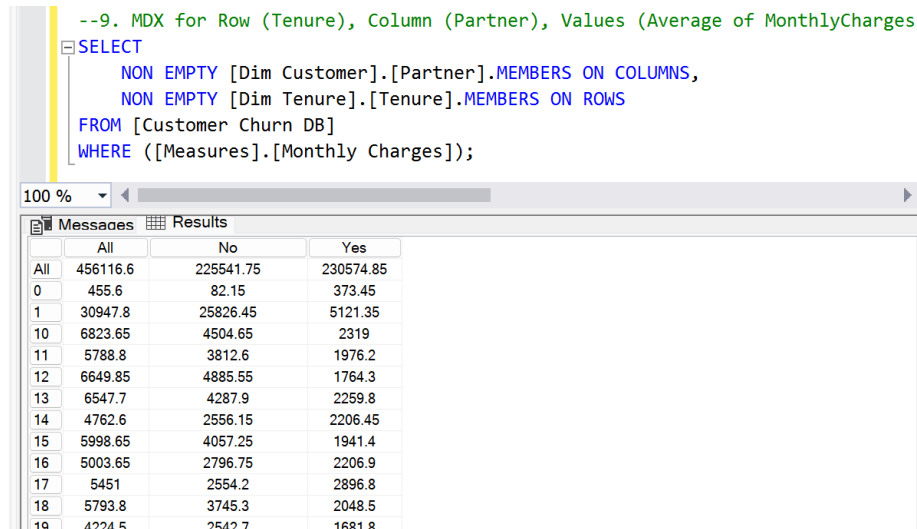

Figure 42: SSAS

Figure 43: MDX

| Partner	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
No	41.08	50.15	56.34	58.65	58.48	60.63	58.49	59.65	60.94	62.73	57.75	58.66	62.64	60.39	62.35	62.42	59.51	55.53	60.41	57.79	60.57	60.41	62.76	63.37	64.05	61.96	65.10
Yes	41.49	52.26	60.62	55.96	54.65	62.01	52.83	59.63	51.47	62.13	61.03	58.12	45.24	59.47	63.04	57.10	66.88	70.65	58.53	57.99	60.89	74.73	60.87	59.96	58.75	56.89	63.82
Total	41.42	50.49	57.21	58.02	57.43	61.00	56.59	59.64	57.25	62.56	58.82	58.47	56.84	60.07	62.67	60.59	62.55	62.66	59.73	57.87	60.71	66.09	61.94	61.56	61.34	59.65	64.65
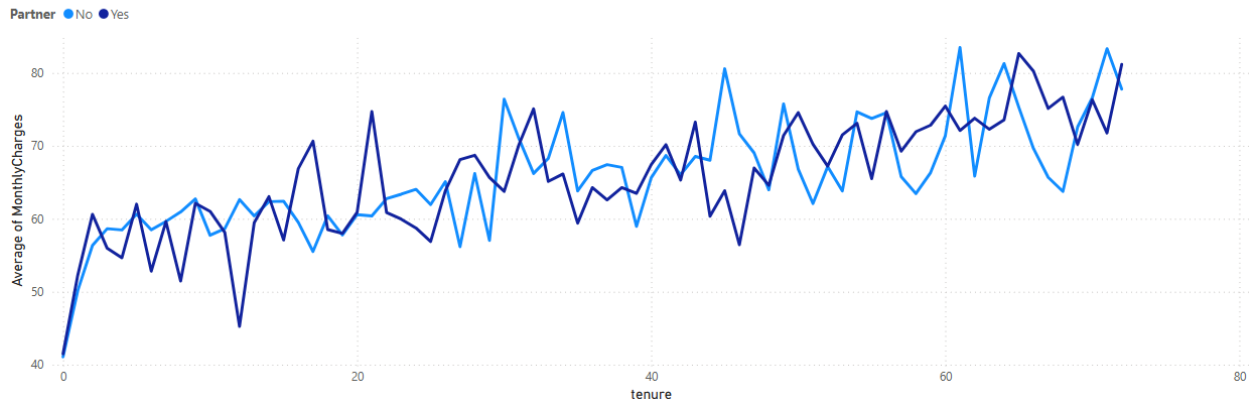
Figure 44: PowerBI
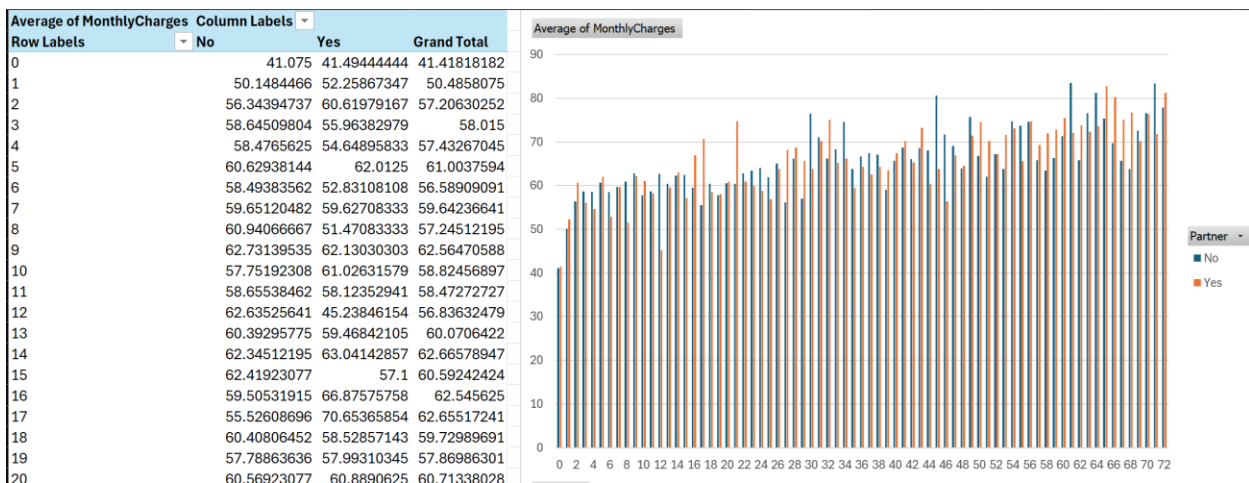
Figure 45: Excel

10. InternetService and PaperlessBilling Customer Counts

Row (InternetService), Column (PaperlessBilling), Values (Count of CustomerID)

Customer experience team checks how many customers per internet service use paperless billing. This helps with launching a sustainability or green billing campaign, promoting eco-friendly practices and boosting retention.

Figure 46: SSAS

```
-- 10. MDX for Row (InternetService), Column (PaperlessBilling), Values (Count
SELECT
    NON EMPTY [Dim Customer].[Paperless Billing].MEMBERS ON COLUMNS,
    NON EMPTY [Dim Internet Service].[Internet Service].MEMBERS ON ROWS
FROM [Customer Churn DB]
WHERE ([Measures].[Customer Churn Fact Count]);
```

	All	No	Yes
All	7043	2872	4171
DSL	2421	1091	1330
Fiber optic	3096	701	2395
No	1526	1080	446

Figure 47: MDX

InternetService	No	Yes	Total
DSL	1091	1330	2421
Fiber optic	701	2395	3096
No	1080	446	1526
Total	2872	4171	7043

Count of customerID by InternetService and PaperlessBilling

PaperlessBilling ●No ●Yes

Figure 48: PowerBI

Count of customerID	Column Labels		
Row Labels	No	Yes	Grand Total
DSL	1091	1330	2421
Fiber optic	701	2395	3096
No	1080	446	1526
Grand Total	2872	4171	7043
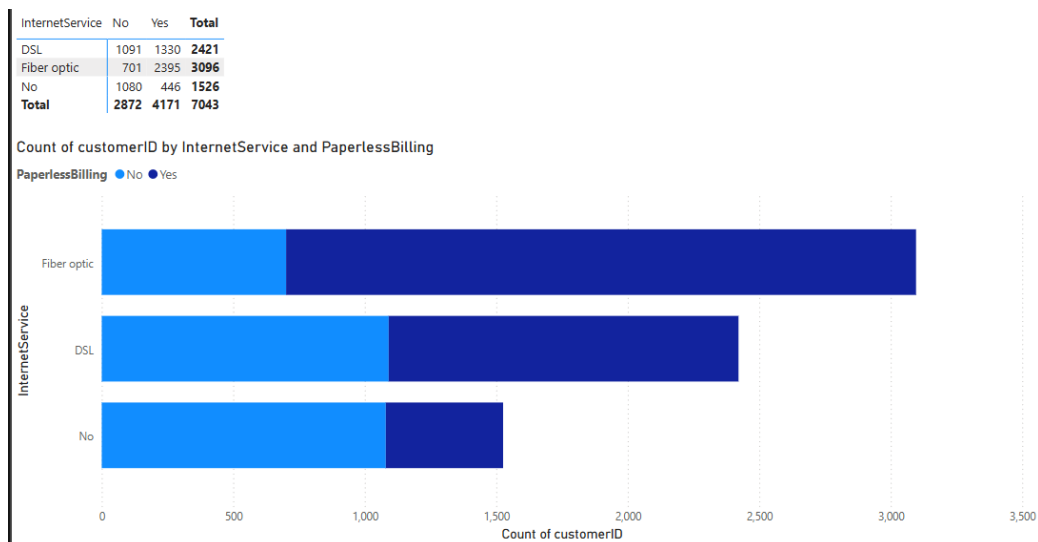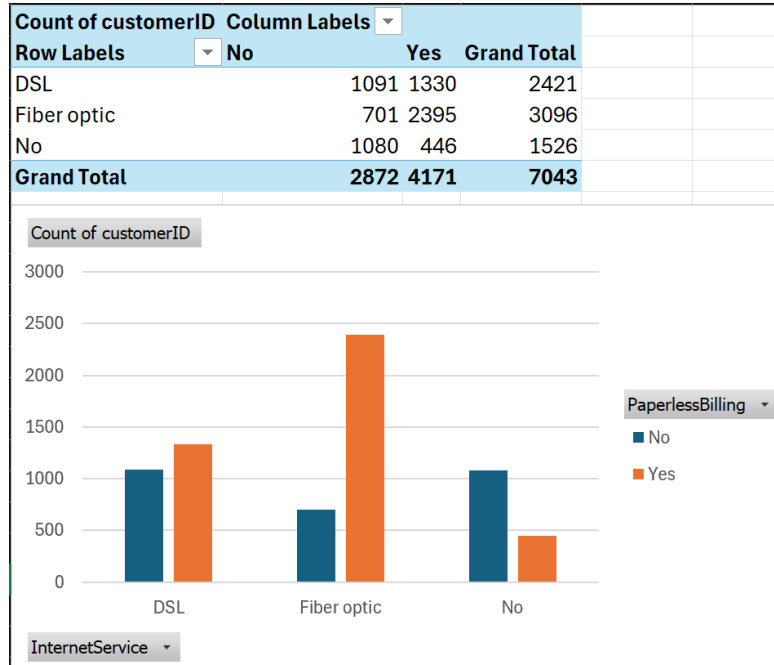
Figure 49: Excel

Table 2: Summary Table

Scenario	BI Goal	Key Value
1. TenureGroup x Churn	Retention campaign	Customer count
2. InternetService x ContractGroup	Pricing discounts	Avg Monthly Charges
3. PaymentCategory x PaperlessBilling	Revenue boost	Total Charges
4. Contract x InternetAvailability	Upgrade strategy	Customer count
5. TenureRange x SeniorCitizen	Senior package	Avg Total Charges
6. InternetService x Dependents	Target singles	Avg Monthly Charges
7. PaymentMethod x Churn	Incentivize auto-pay	Customer count
8. ContractGroup x Gender	Gender-based promotion	Total Charges
9. Tenure x Partner	Loyalty program	Avg Monthly Charges
10. InternetService x PaperlessBilling	Paperless campaign	Customer count

III. Data mining (2 algorithms, deep learning)

Overview: In the part, we will use 6 algorithms to predict Customer Churn in order to decide which model is the best for classification of Churn customer values. From the model, we will extract the rules for this dataset.

We will define which features are the most important to Churn values. We use Correlation matrix, Random Forest algorithm, Permutation on Test set to figure it out.
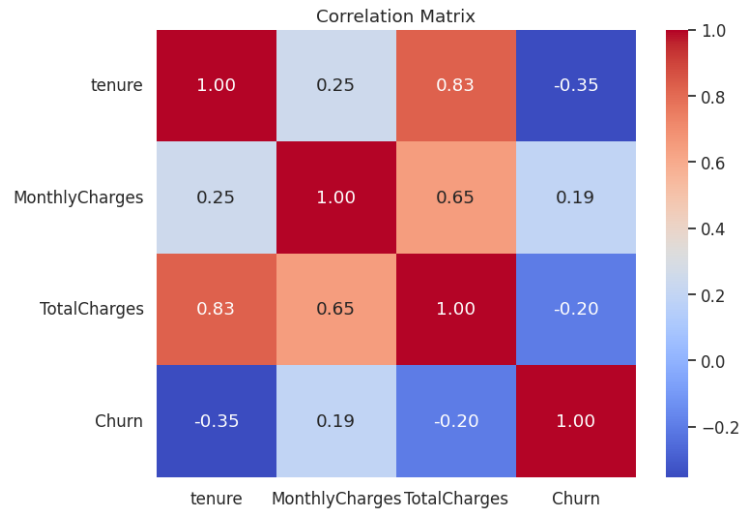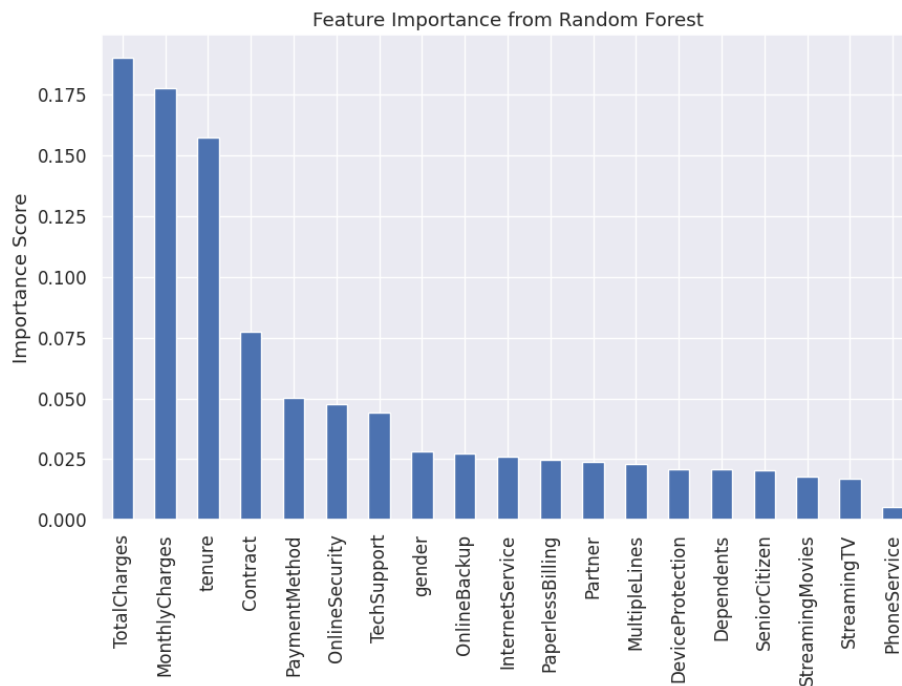

Figure 50: Correlation Matrix

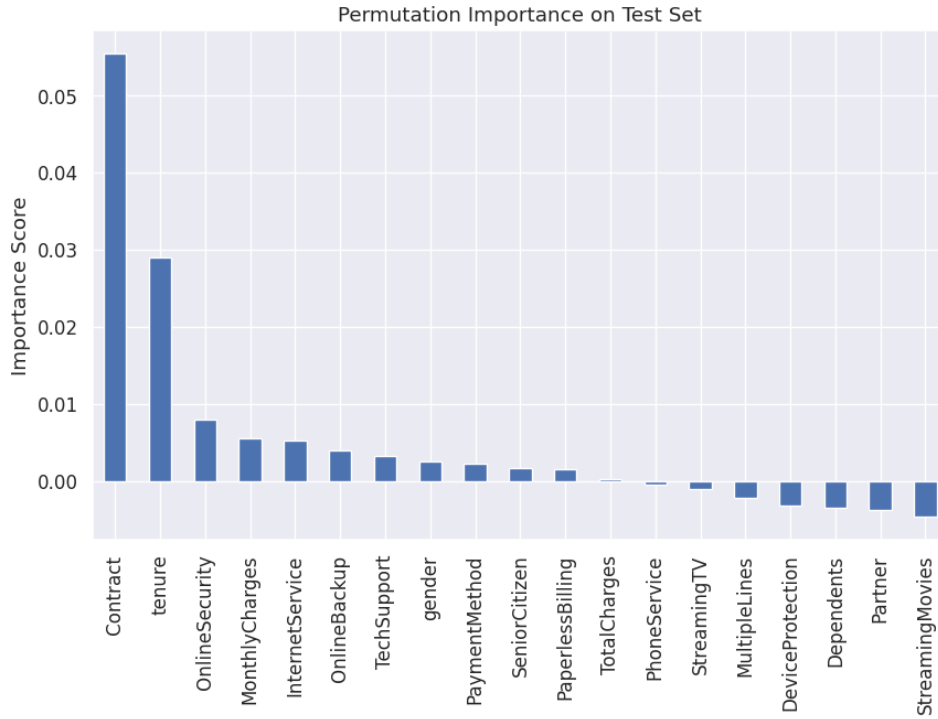*Figure 51: Feature importance by Random Forest*

Figure 52: Permutation importance on Test set

After feature engineering, we can notice that Contract, Tenure, and Monthly Charges are 3 most important features affecting Churn outcomes.
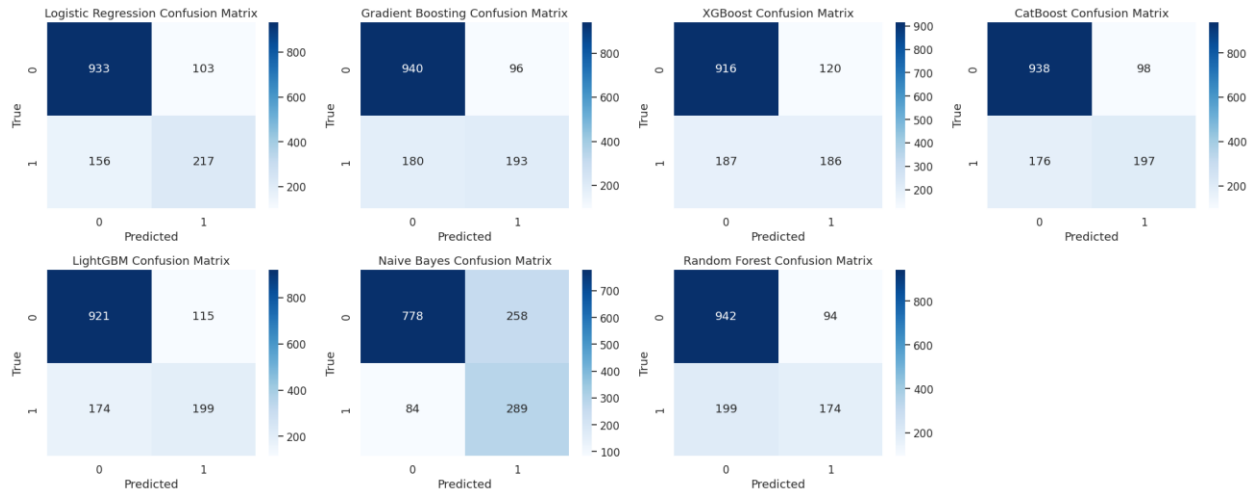

Figure 53: Confusion Matrixes

We use ROC Curves to evaluate all Models. The results show that Logistic Regression, Gradient Boosting, CatBoost are the highest accuracy models.
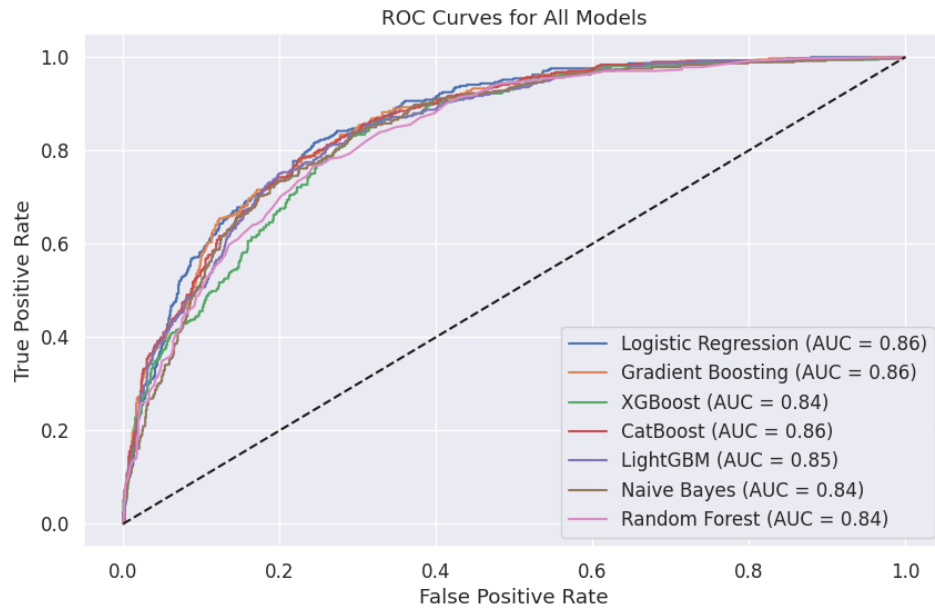

Figure 54: ROC curves

After applying all models to predict the Churn values. We use Radar char to show important features of each model.

Figure 55: Top Feature radar charts

We print out the Churn values (yes = 1, no = 0) of the actual, Predicted **Logistic Regression**, and predicted **Gradient Boosting** values to compare it together.

```
dtype: float64
      Actual  Predicted Logistic  Predicted Gradient
185        1                    1                   1
2715       0                    0                   0
3825       0                    0                   0
1807       1                    1                   1
132        0                    0                   0
1263       1                    0                   0
3732       0                    0                   0
1672       0                    0                   0
811        1                    0                   0
2526       1                    0                   0
```

Figure 56: Predicted values comparison

Lastly, we define Rules for churn classification. The result below is the rule with Scaled values.

```
Classification Rules for Churn (Scaled):
|--- Contract <= 0.50
|   |--- MonthlyCharges <= 0.12
|   |   |--- tenure <= -1.18
|   |   |   |--- class: 0
|   |   |--- tenure >  -1.18
|   |   |   |--- class: 0
|   |--- MonthlyCharges >  0.12
|   |   |--- tenure <= -0.77
|   |   |   |--- class: 1
|   |   |--- tenure >  -0.77
|   |   |   |--- class: 0
|--- Contract >  0.50
|   |--- MonthlyCharges <= 0.96
|   |   |--- Contract <= 1.50
|   |   |   |--- class: 0
|   |   |--- Contract >  1.50
|   |   |   |--- class: 0
|   |--- MonthlyCharges >  0.96
|   |   |--- Contract <= 1.50
|   |   |   |--- class: 0
|   |   |--- Contract >  1.50
|   |   |   |--- class: 0
```

Figure 57: Classification Rules for Churn (Scaled)

We transform again the rules from scaled values to original values which are more meaningful in Business intelligence decision.

```
Classification Rules for Churn (Original Values):
|--- Contract is Month-to-month
 |--- MonthlyCharges <= 68.63
  |--- tenure <= 3.00
   |--- class: No Churn
   |--- class: No Churn
  |--- tenure <= 14.00
   |--- class: Churn
   |--- class: No Churn
 |--- MonthlyCharges <= 93.68
  |--- Contract is Month-to-month or One year
   |--- class: No Churn
   |--- class: No Churn
  |--- Contract is Month-to-month or One year
   |--- class: No Churn
   |--- class: No Churn
```

Figure 58: Classification Rules for Churn (Original Values)

Explanation of Classification Rules

1. Contract is Month-to-month:

- The rule only applies to customers with a Month-to-month contract. These customers are known to be more prone to churn compared to long-term contract holders.

2. If MonthlyCharges <= 68.63:

Then check tenure:

- If tenure <= 3 → **No Churn**

Customers with low charges and very new are predicted not to churn.

- If tenure <= 14 → **Churn**

If the customer is moderately new (3 to 14 months), the model predicts churn.

Else → **No Churn**

If tenure is longer than 14 months, churn risk drops again.

3. If MonthlyCharges <= 93.68 (and more than 68.63):

Regardless of tenure, if **Contract is Month-to-month or One year**:

→ **No Churn**

These customers pay moderately high charges but are still likely to stay.

4. If MonthlyCharges > 93.68:

Again, if Contract is Month-to-month or One year:

→ **No Churn**

Even high-paying customers are predicted not to churn, possibly because they're receiving more services and are satisfied.

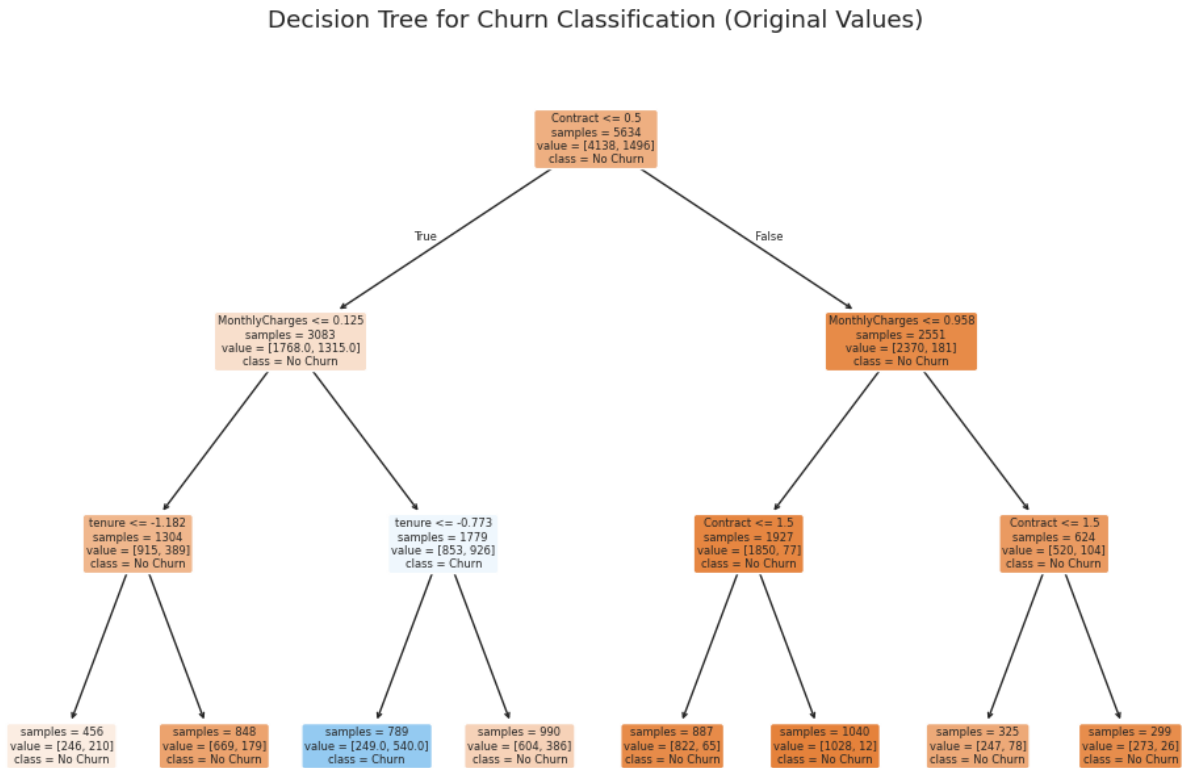We show the Decision tree for Churn classification.

Figure 59: Churn decision tree

We also add a Grouped bar char about Average Monthly Charges by Tenure bins and Churn values.
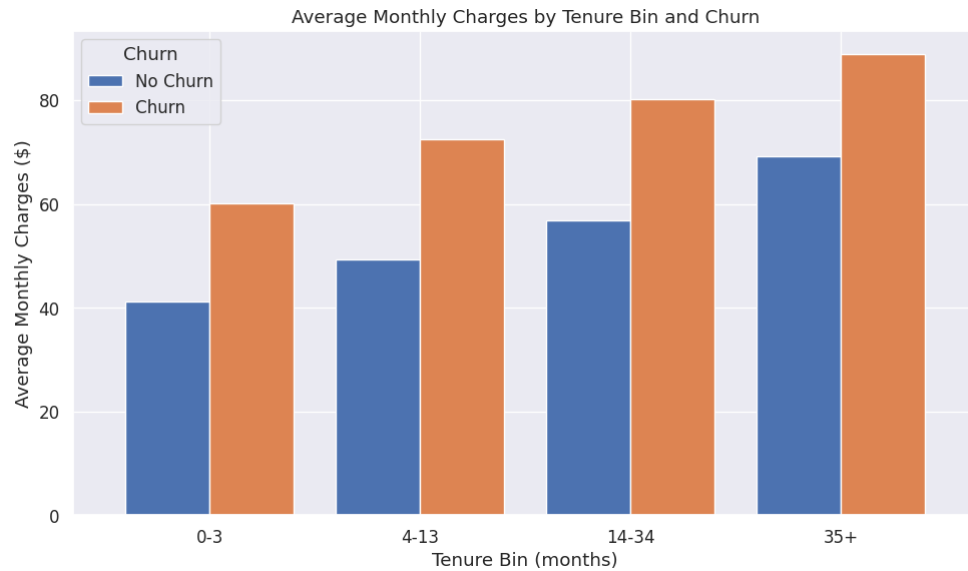

Figure 60: Average Monthly Charges by Tenure Bin and Churn

Business Rule from Data mining:

Customer Condition	Churn
Month-to-month, **high** MonthlyCharges, **low** tenure	Churn
Month-to-month, **low** MonthlyCharges	No Churn
One-year or Two-year contract	No Churn

Conclusion: Customers on **monthly contracts**, paying **high monthly fees**, and **new to the service (low tenure) → high** likelihood of churn.

END