IST- 722 Data Warehouse Project

FudgeMart Inc. Data Warehouse & B.I solution

Project by-

Tanya Shrivastava

Vijay Bhat

Shashank Srihari

Palaniappan Muthukkarappan

Table of Contents

1. OVERVIEW SECTION 3
1.1. COMPANY OVERVIEW 3
1.2. BUSINESS CASE 3
1.3. SCOPE 3
1.4. FUNCTIONAL REQUIREMENT 4
1.5. TEAM MEMBERS AND ROLES 4
1.6. TASKS PERFORMED 5
1.7. KEY STAKEHOLDERS 5
2. ANALYSIS SECTION7
2.1. BUSINESS PROCESSES 7
2.2. BUS MATRIX&HIGH-LEVEL DESIGN 8
2.3. DETAILED DIMENSIONAL DESIGN WORKSHEET 9
3. DETAILED DESIGN SECTION
3.1. DIMENSIONAL HIERARCHIES 12
3.2. DIMENSIONAL MODEL 14
3.3. DETAILED ETL FLOW FOR EACH SOURCE TO TARGET 16
4. BUSINESS INTELLIGENCE: PowerBI 26

1. OVERVIEW

1.1 Company Overview:

We have been presented with FudgeFlix, consisting of two subsidiary companies, Fudgemart and Fudgeflix.

Fudgemart- It is an online retailer with 6 departments, 53 primary products (such as such as hardware, electronics, sporting goods etc.) being sold across them with employees working in different departments.

Fudgeflix- It is an online DVD by mail and video on demand service providing users with numerous movies across different genres.

1.2 Business Case:

The Data Warehouse is being created in order to enable the seamless integration of the two given databases into a single data warehouse, compatible with Business Intelligence tools. Essentially, this is done in order to achieve Subject-orientation, Non-volatility, Integration, and Time- Variance, hence fulfilling the very need for a Data Warehouse.

Currently FudgeMart and FudgeFlix have two independent databases along with a third supporting database. FudgeFlix Corp. has commissioned the creation of a centrally integrated DW/BI solution from scratch, which will involve:

- Staging the required data
- Performing ETL

• Transferring the data onto the warehouse

• Using BI tools to generate actionable insights

1.3 Scope:

The scope of this project is defined around the following deliverables:

• Delivery of the data definition for the ODS and the data warehouse that is responsive towards the KPI and reporting requirements of the organization, including designed formulas to calculate values from the data for the required KPIs.

• Accepted and feasible technical architecture for RDBMS, ODS and DW

• Delivery of the defined format of the data warehouse, data marts and selection of BI tools

• Test and Acceptance of the Data warehouse

• Define processes for maintaining the DWH (frequency of uploading data, data quality evaluation, reporting, DB maintenance)

• Design a training and usage plan.

1.4 Functional Requirement:

Simple requirement

- Analyzing Product Review Stars
- Movie Popularity

Complex requirement

- Aggregated Sales Reporting
- Customer Promotional Strategy

1.5 Team Members and Roles:

The following table summarizes the responsibilities of each team member as their defined roles in the project.

Team Member	Role
Vijay Bhat	**Project Management**
Tanya Shrivastava	**ETL Specialist**
Palani	**Data Engineer**
Shashank Srihari	**BI Developer**

1.6 Tasks Performed:

The various tasks performed in order to accomplish the project are as follows:

• Creation of High-level dimensional model

• Creation of Detailed level dimensional model

• Execution of ETL in SSIS

1. Staging the data for simple and complex process.
2. Transitioning to FactTables.

• Creation MOLAP cubes using SSAS and using PowerBI (and Tableau) to generate dashboards

1.7 Key Stakeholders:

The key stakeholders of the Fudge Mart are:

Executives –The executive team needs to be clear on the importance of B2B e commerce to set the direction for the company's mission and thereby achieve strategic targets.

Marketing – The marketing department is known for being your organization's branding police and will have a considerable number of requirements for a successful B2B e commerce site from the perspective of marketing. Without involving a key player from the marketing department on your ecommerce team, you may end up with an ecommerce site that is limited in how it can promote and sell your products or one that is not visually appealing to your customers.

Sales – A leader from your sales department must be included on any successful ecommerce team to ensure that your website accurately reflects the sales arrangements offered to your customers in one-on-one personal sales. For example, if certain customers are offered specific negotiated pricing for the products that they purchase from you, your site will need to reflect that pricing arrangement.

IT – Ecommerce is often mistakenly assumed to be an IT project because the business transaction happens on a computer. While IT will play a role in making sure that your ecommerce system integrates properly with your ERP and other back-end systems, the implementation and day-to-day management of an enterprise B2B e commerce should rarely fall solely on the backs of the IT group.

Operations – Having a member of your operations team on the e-commerce launch team will help ensure that you consider all of the ways e-commerce can generate efficiencies for your organization as well as what operational hurdles you will need to watch out for.

Apart from these, other important key stake holders are:

• Database Administrators
• Data Architect
• End Users (registered and guest users)
• ETL Specialist
• Project Manager
• Data Analyst
• BI Engineer

2. ANALYSIS SECTION

2.1 Business Processes:

- **Product Review Stars:**

1. Analyzing Products reviewed by Customers based on the star rating on the scale of 1-5. These are drilled down by month of sale, city of customer, category of the product and region sold.
2. FudgeMart offers a system in which FudgeMart customers can review FudgeMart products
3. Product reviews help us track the quality of products over time
4. Transaction

- **Aggregated Sales Reporting:**
  1. Aggregating Customers from Fudgemart and Fudgeflix and analyzing sales by region, month, the day of the week. Highest selling products are also identified.
  2. FudgeMart sells products to customers in an online store
  3. FudgeFlix has customers subscribed to a monthly payment plan
  4. Tracking the total sales for the company allows us to track the growth of company
  5. Transaction

2.2 Bus Matrix & High-Level Design

Instructions!													
Business Process Name	**Fact Table**	**Facts**	**OrderDate**	**Product**	**DimCustomer**	**MovieTitle**	**ReviewDate**	**ShippedDate**	**Shipper**				
Total Sales	FactTotalSales	Sold Amount, quantity, unit price	X	X	X								
Movie Orders	FactMovieOrders	Count, Method, Rating	X		X	X							
Customer Reviews	FactFudgemartReviews	ReviewStars		X	X		X						
Order Fulfillment	FactOrderFulfillment	DaysBetweenOrderAndShipped	X		X			X	X				

Detailed Dimensional Design Worksheet:

FactTotalSales (Complex Business Process)

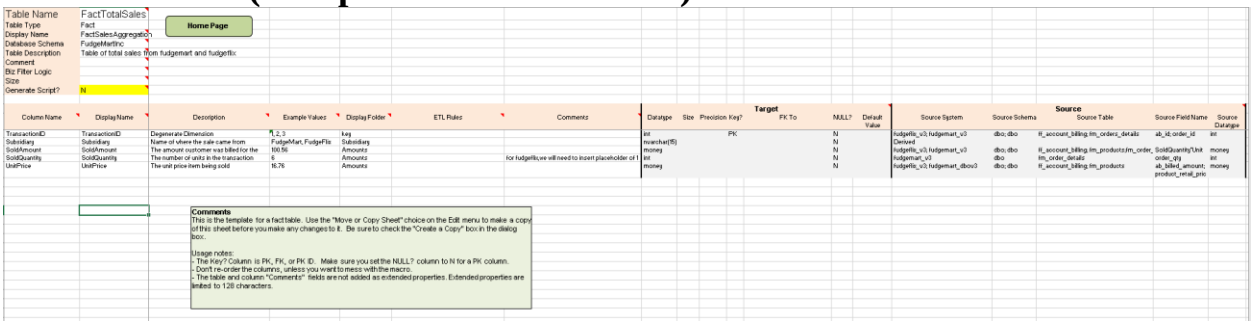


Table Name	FactTotalSales
Table Type	Fact
Display Name	FactSalesAggregation
Database Schema	FudgeMartInc
Table Description	Table of total sales from fudgemart and fudgeflix
Comment	
Biz Filter Logic	
Size	
Generate Script?	N

Home Page

							Target						Source				
Column Name	Display Name	Description	Example Values	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
TransactionID	TransactionID	Degenerate Dimension	1,2,3	Key			int			PK		N		fudgeflix_v3.fudgemart_v3	dbo; dbo	ft_account_billing.tm_orders_details	ab_id; order_id	int
Subsidiary	Subsidiary	Name of where the sale came from	FudgeMart, FudgeFlix	Subsidiary			nvarchar(15)					N		Derived				
SoldAmount	SoldAmount	The amount customer was billed for the	100.56	Amounts			money					N		fudgeflix_v3.fudgemart_v3	dbo; dbo	ft_account_billing.tm_products;tm_order_	SoldQuantity*Unit	money
SoldQuantity	SoldQuantity	The number of units in the transaction	5	Amounts		for fudgeflix,we will need to insert placeholder of 1	int					N		fudgemart_v3	dbo	tm_order_details	order_qty	int
UnitPrice	UnitPrice	The unit price item being sold	16.76	Amounts			money					N		fudgeflix_v3.fudgemart_dbo3	dbo; dbo	ft_account_billing.tm_products	ab_billed_amount; product_retail_pric	money

Comments
This is the template for a fact table. Use the "Move or Copy Sheet" choice on the Edit menu to make a copy of this sheet before you make any changes to it. Be sure to check the "Create a Copy" box in the dialog box.

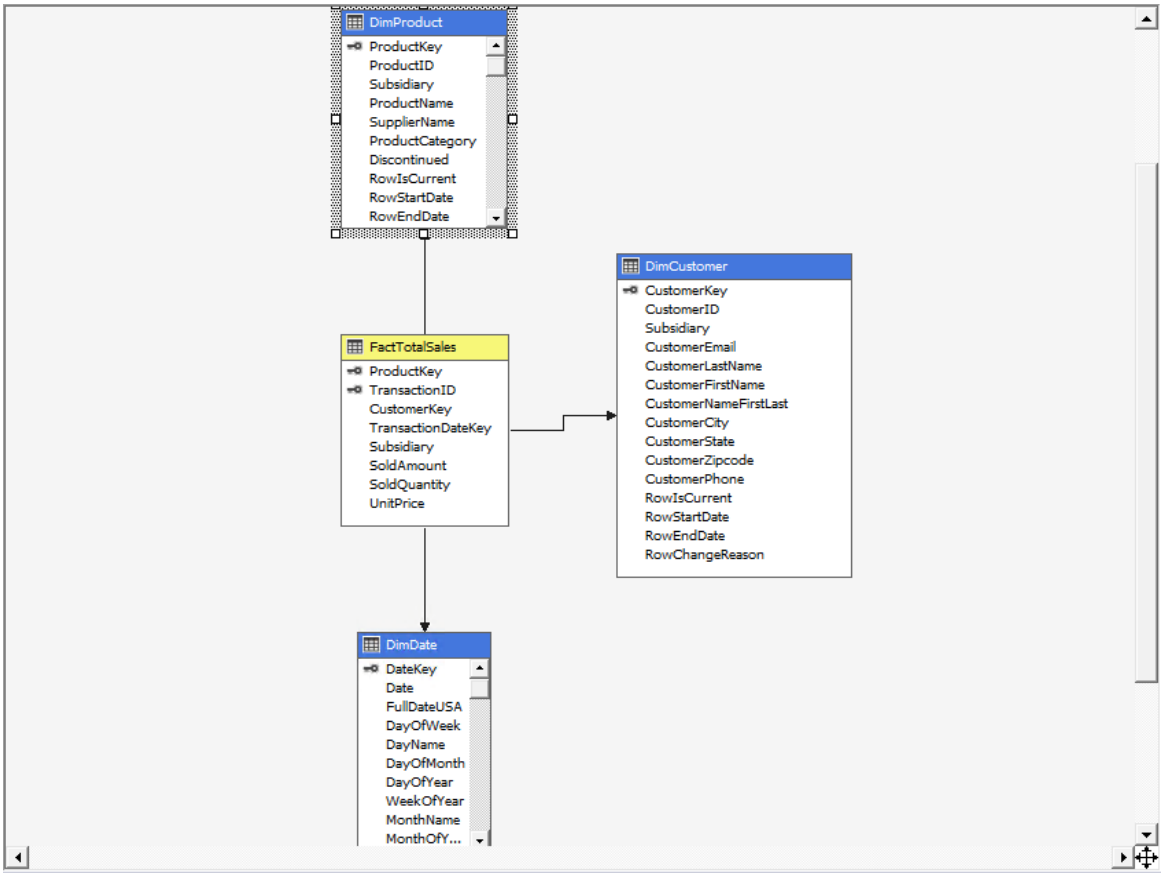
Usage notes:
- The Key? Column is PK, FK, or PK ID. Make sure you set the NULL? column to N for a PK column.
- Don't re-order the columns, unless you want to mess with the macro.
- The table and column "Comments" fields are not added as extended properties. Extended properties are limited to 128 characters.

FactFudgeMart Reviews (Simple Business Process)

Table Name	FactFudgemartReviews
Table Type	Fact
Display Name	FactStarReviews
Database Schema	FudgeMartInc
Table Description	Table on star reviews from fudgemart
Comment	
Biz Filter Logic	
Size	
Generate Script?	N

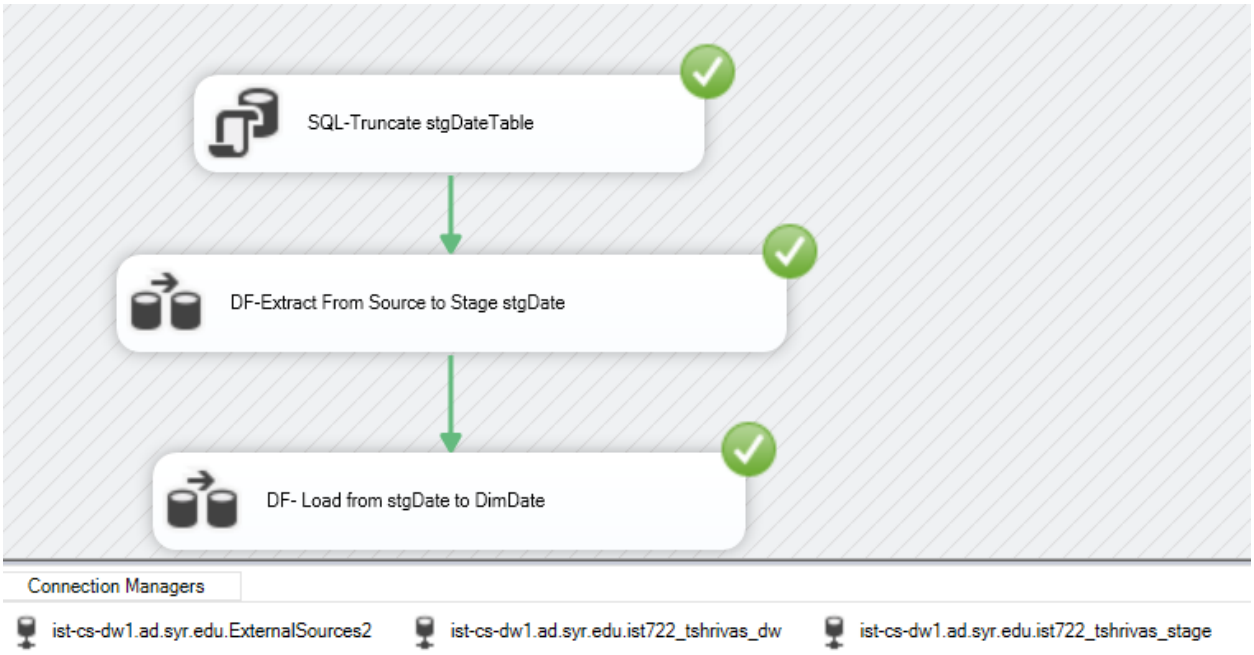
		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
ProductKey	ProductKey	int			FK,PK	DimProduct.Product	N	ETL Process					
CustomerKey	CustomerKey	int			FK,PK	DimCustomer.Custo	N	ETL Process					
ReviewDateKey	ReviewDateKey	int			FK	DimDate.DateKey	N	ETL Process	Fudgemart	dbo	customer_product_review_date		datetime
ReviewStars	ReviewStars	int					N	ETL Process	Fudgemart	dbo	customer_product_review_stars		int

Dimensional Model (Star Schema):

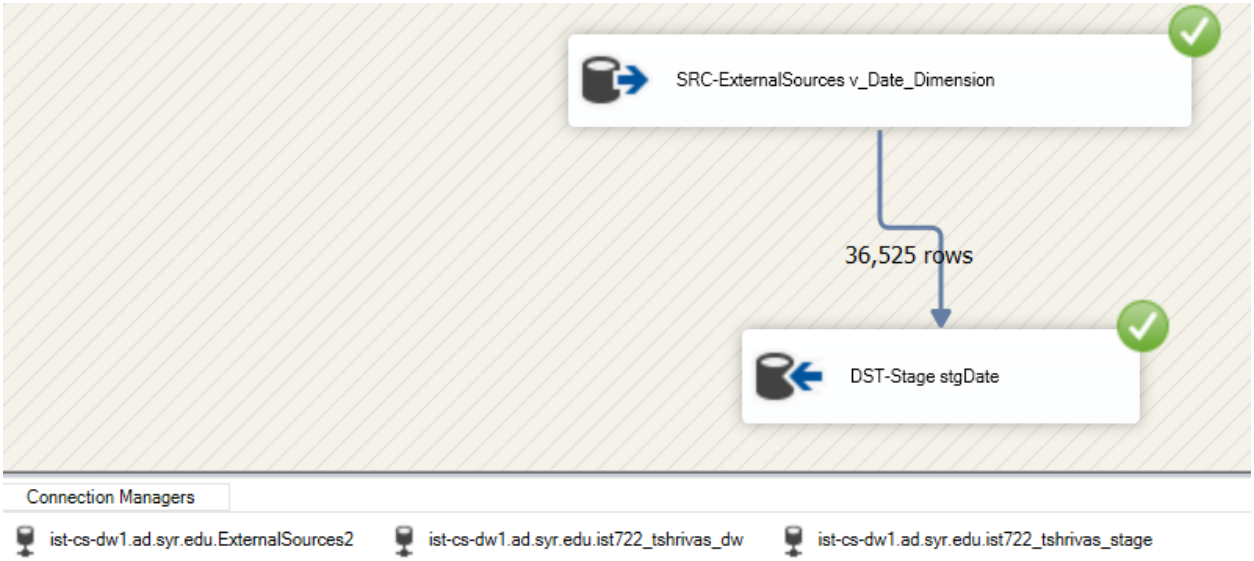


ETL Steps

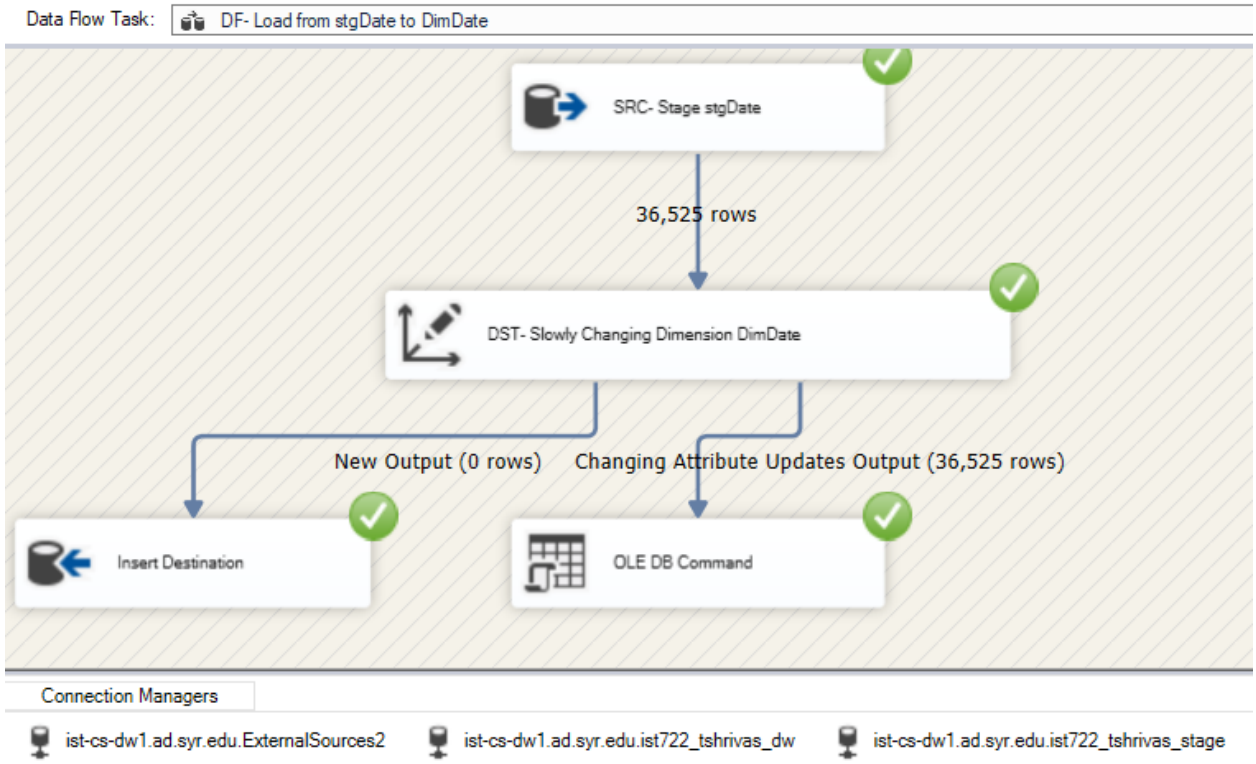
External Sources -> Stage -> DW.



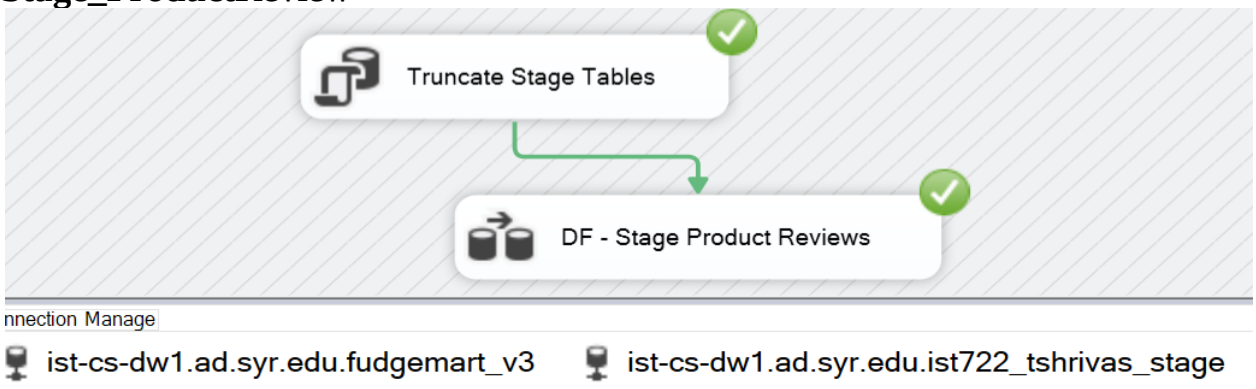
DF-Extract From Source to Stage stgDate



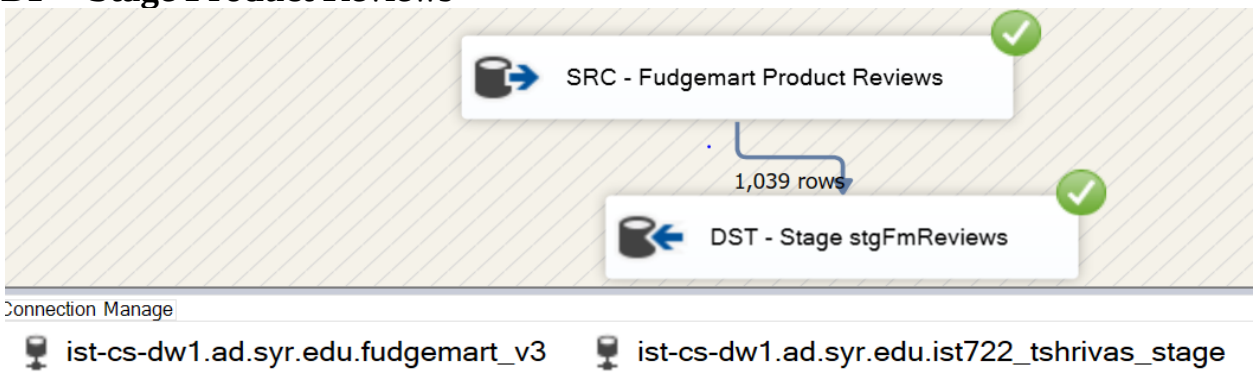
DF-Load from stgDate to DimDate



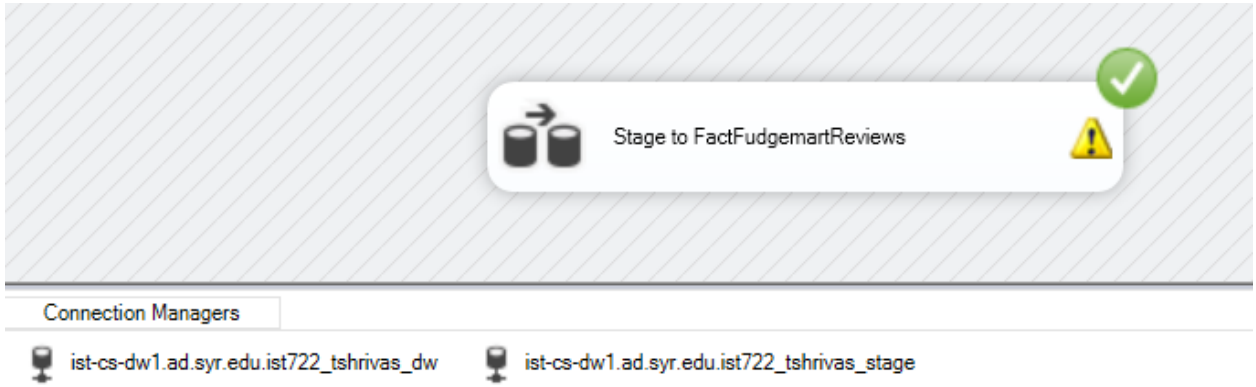
Stage_ProductReview



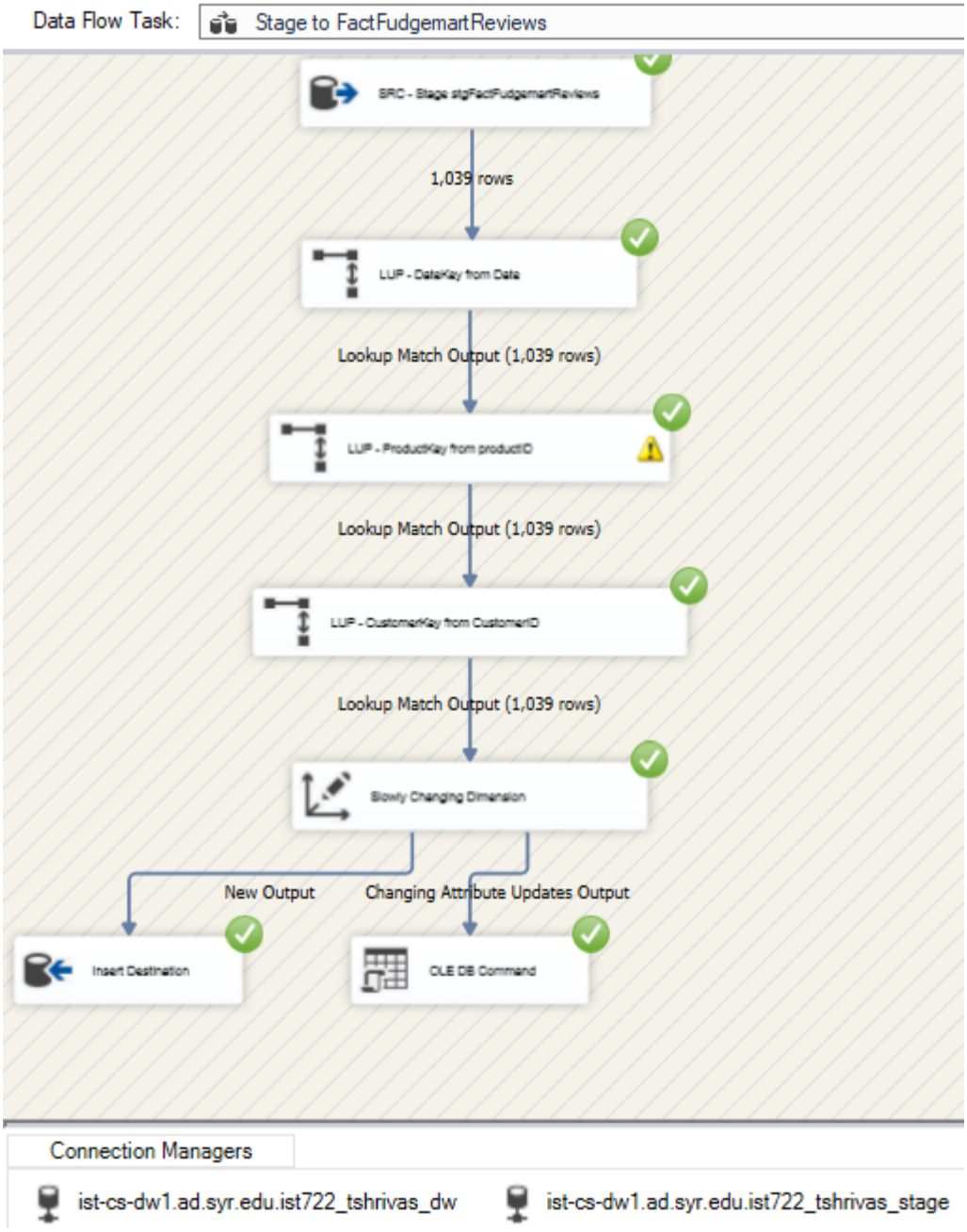
DF – Stage Product Reviews



DW_FudgemartReview

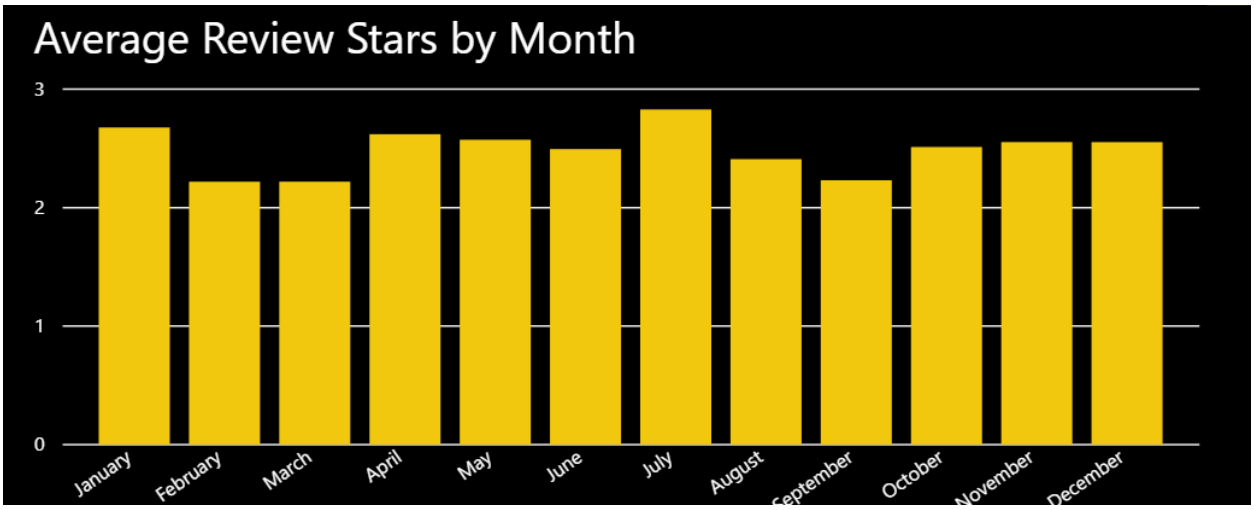


Stage to FactFudgemartReviews



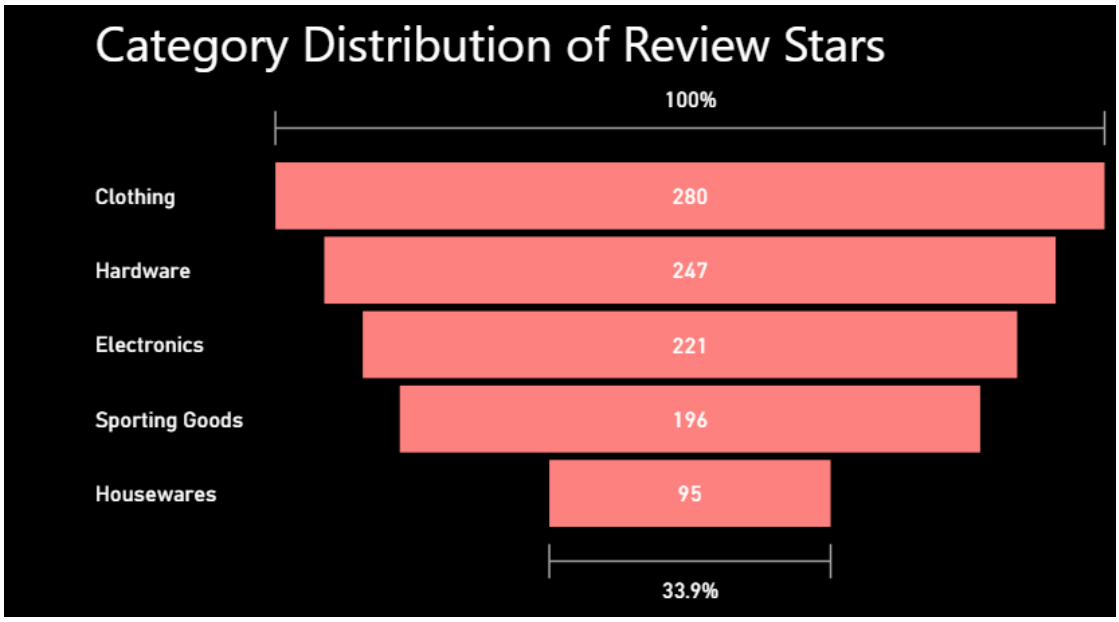
3 BUSINESS INTELLIGENCE: PowerBI

Star rating vs Month graph.

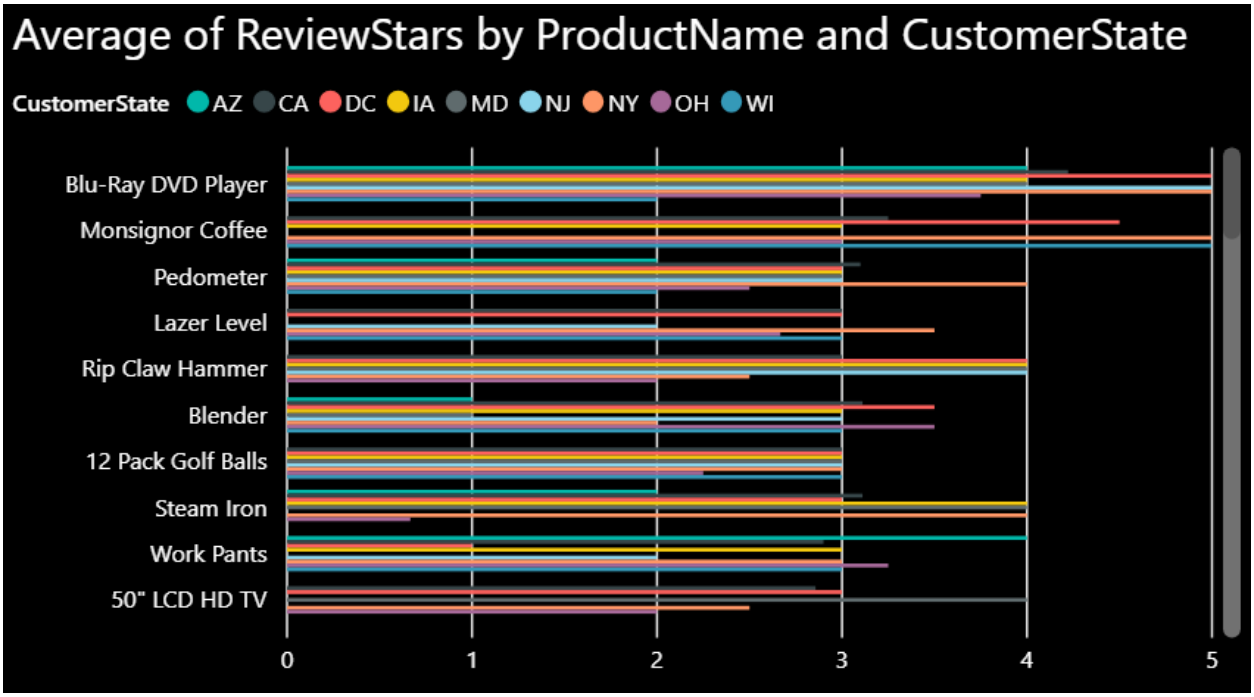


This graph shows us the average star rating for each month. We can see that star ratings increase during July and January.

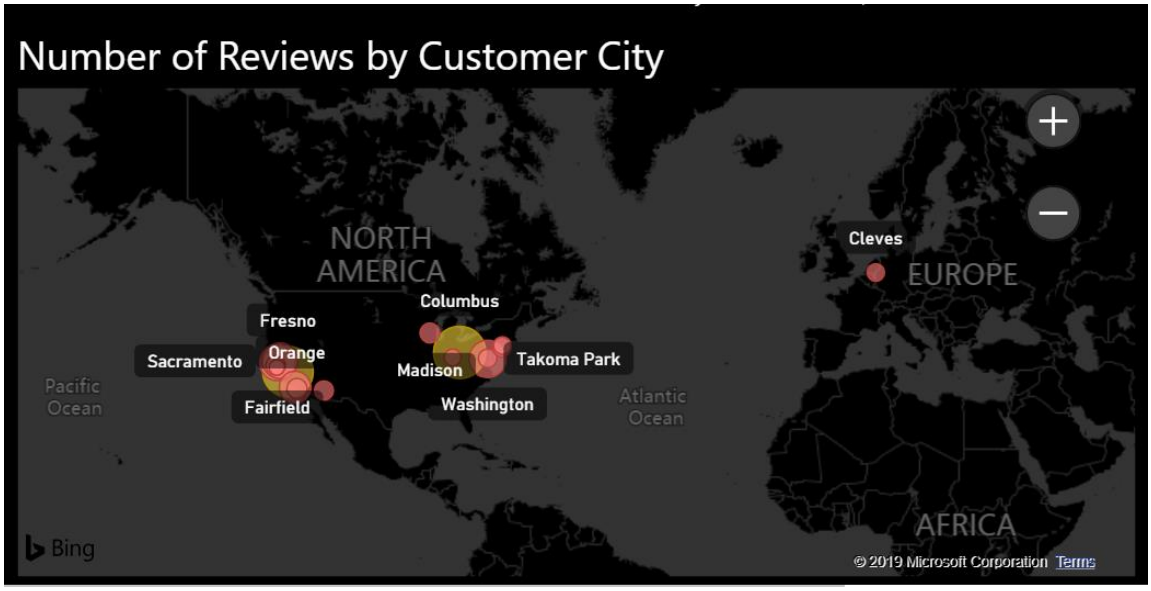
Product category vs Number of reviews



Star rating of products with respect to the states.

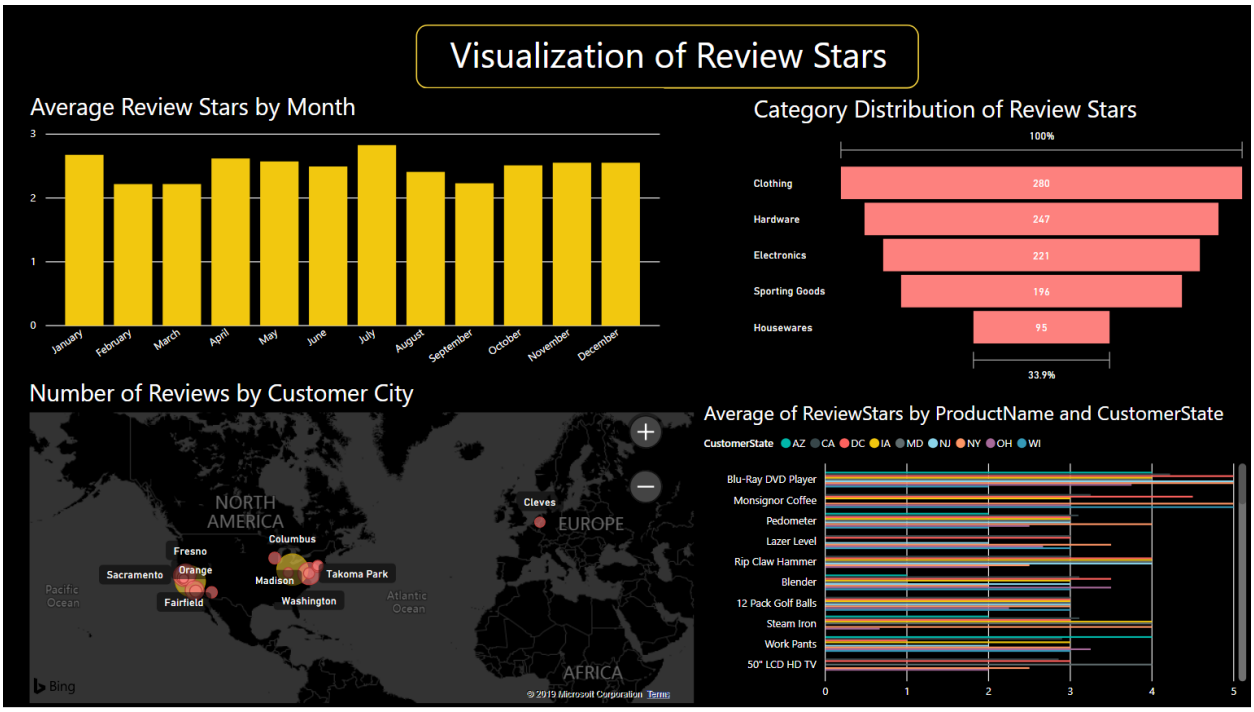


This graph shows the sale of different products in the respective states. The analysis of what products are popular in which states and regions gives an idea to the business as to which promotions and support marketing operations.

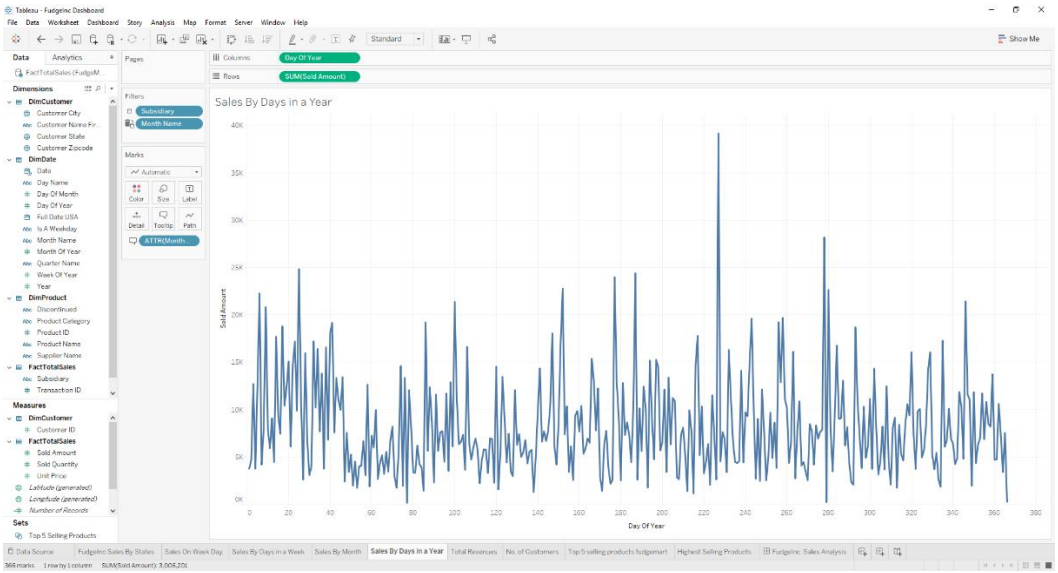


This graph shows which cities have the most responsive customers. This helps the business to identify key concerns raised by consumers and to address these concerns on a very scaled-up level.

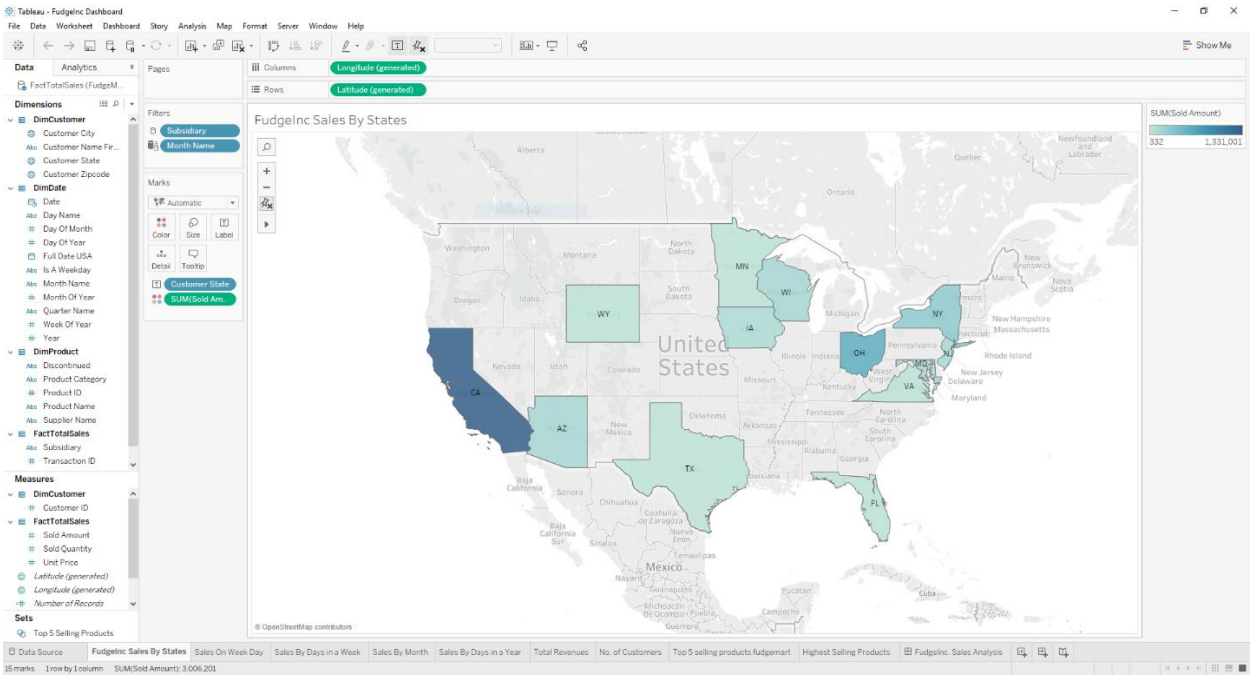
Overall visualization on PowerBI



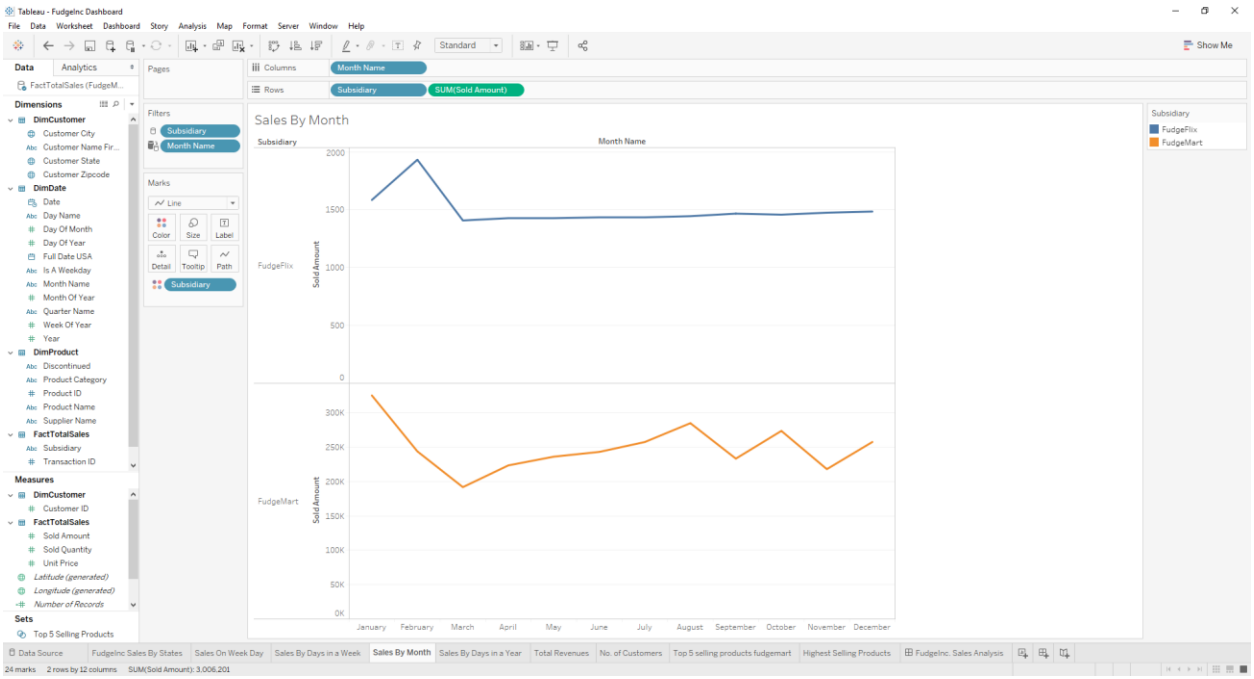
Data Story: The overall dashboard for the simple process we chose to implement is as depicted above. The dashboard contains all the graphs that we previously explained and gives an overall view to the business users.



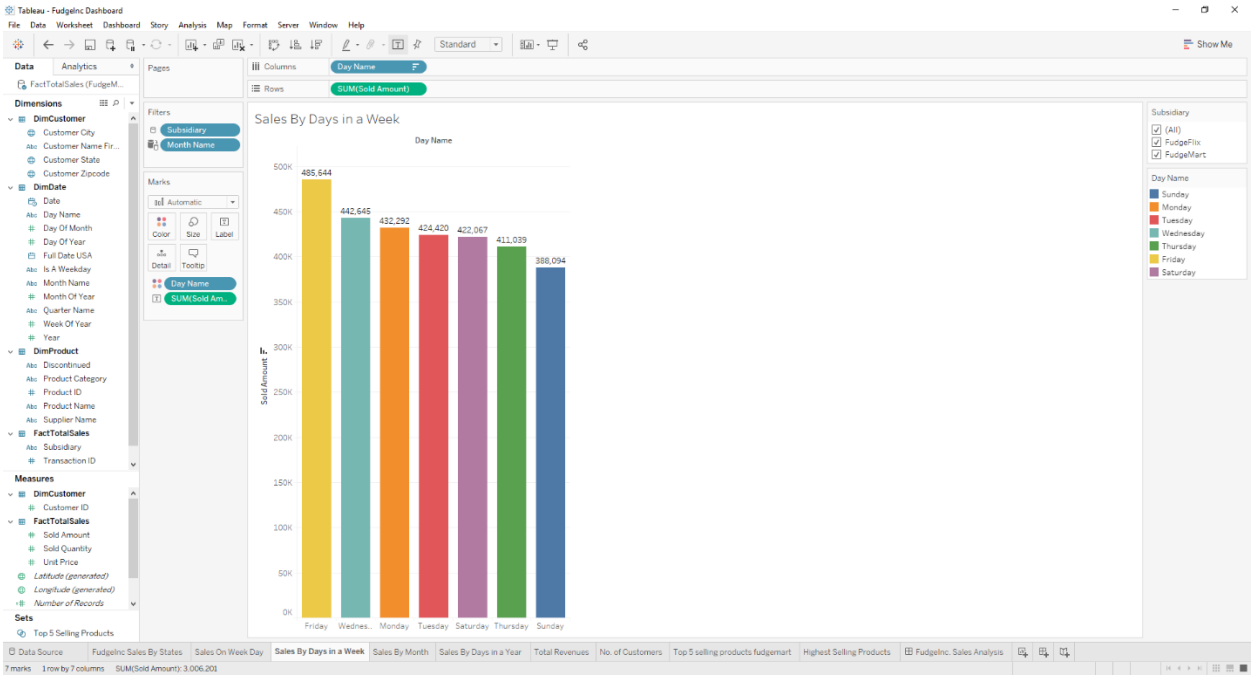
The above graph shows the sales by day for every day in the year. Drilled own, this shows us spikes in sales on particular days and when sales die down. This helps the business to identify particular days in the year to run promotional offers to boost sales.



The above graph shows the sales by state and helps the business identify regions of sale and shows where the marketing strategies can be focused on.

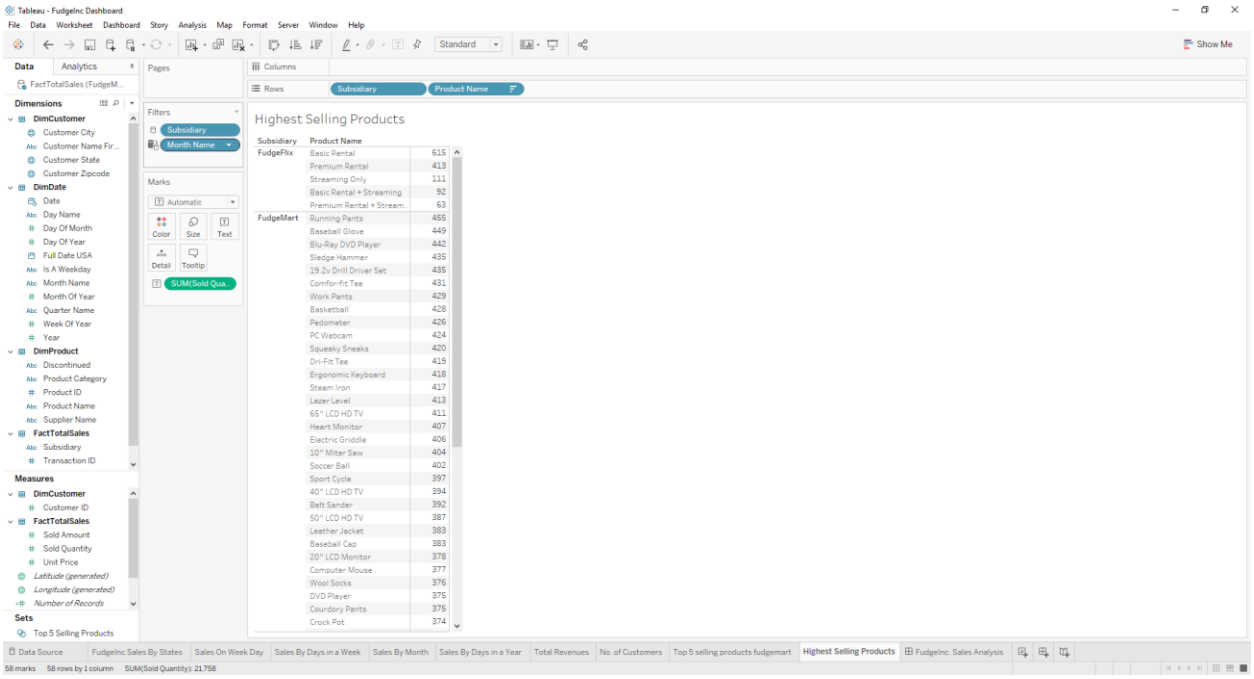


This graph shows the sales for FudgeMart and FudgeFlix by month. We see spikes at different times of the year for each. The business should be able to identify when the promotional offers can be used to boost sales.

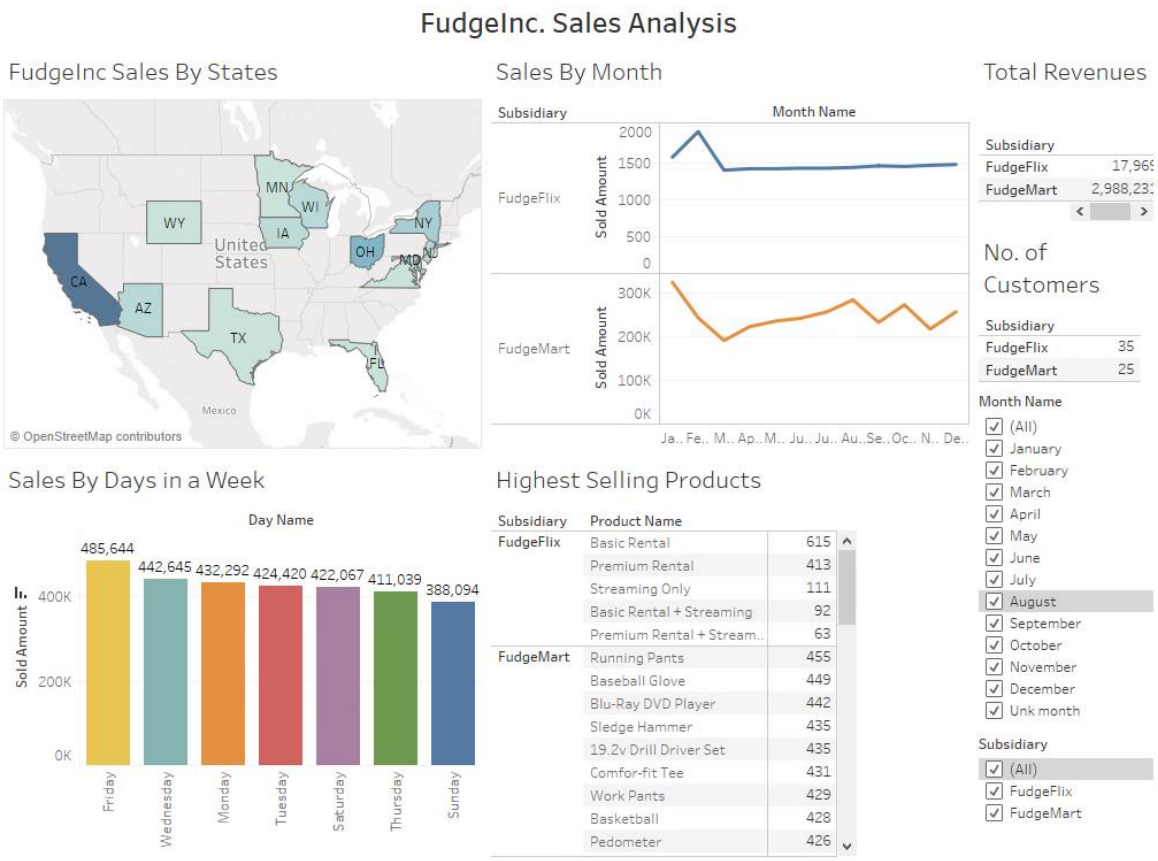


This graph shows the sales for both FudgeMart and FudgeFlix by day of the week. We see the spikes occurring in the days around the weekend and the business culd

use this information to focus their marketing strategies to take place on days of lower sales or just before the weekend.



The highest selling products metrics is essential in order to understand what products are hot at the moment and what products to promote and market. The business could use this information to also determine patterns for buying customers.



This last graph shows the aggregated view of each of the graphs explained above and shows the dashboard view of the complex process we implemented for our project.