Business Intelligence II Assignment

Managing Cubes, Reports and Dashboards

University:	New University of Lisbon	
Faculty:	Information Management School	
Authors:	Jan-Benedikt Jagusch	M2016022
	Jacob Hastrup Madsen	M2016087
	Marcel Motta Do Nascimento	M2016337
Course:	Business Intelligence II	
Professor:	Miguel Neto	
	Rui Monteiro	
	Luis Pedro Batista	
	Miguel Gago	
Date:	Saturday, June 10, 2017	

1 Abstract

This paper deals with the analytical and reporting perspective of business intelligence, applied to a fictional business scenario of IMSports. For analytical purposes, an Online Analytical Processing (OLAP) Cube is built on top of a data warehouse, allowing users to partition data, create key performance indicators and aggregate measures across dimensions. For reporting purposes, a set of reports are created, while dashboards are built to deliver more insightful reporting.

The tools are built around the philosophy of the management tool Balanced Scorecard (BSC), allowing management to keep track of business performance from the classic four perspectives: financial, customer, internal business process and learning and growth.

Keywords: OLAP Cube, Balanced Scorecard, Reporting, KPI, Dashboards

Contents

2 Presentation of Business Scenario

IMSports is a fictional bicycle manufacturer with focus on metal and composite bicycles. Operating from Washington, the organization produces in the USA and Mexico and ships to America, Europe and Asia, where it has specific regional sales teams. The company sells to private customers, buying from the online store, as well as commercial customers buying in physical stores.

In the previous year, IMSports created a revenue of 43 million €, yielding in an excellent annual growth of 30% and a profitability of 32%. Around 70% is generated locally in North America (USA, Canada), 20% in Europe (France, Germany, UK) and 10% in Australia. A big share of the revenue (75%) falls from road bikes and mountain bikes. 26% of revenue are created over the online retail.

IMSports is expanding, acquiring competitors and entering new markets, and hence their business complexity increases while management goes from a daily operational focus to a more long-term strategy. Thus, management requires more detailed, precise and insightful information, for IMSports to maintain competitive advantage.

IMSports already has a data warehouse for all sales transactions, leaving behind the concerns of data storage, but still lacks adequate analytics and reporting tools, making business performance difficult to measure.

Thus, management is now keen on approaching business performance with a balanced scorecard, making sure that the organization knows its goals while being able to follow-up on indicators. For the implementation of balanced scorecard, management is willing to allocate resources for the necessary analytical and reporting tools and hence making the BI solution complete.

2.1 About Balanced Scorecard

Balanced Scorecard is a business performance management tool with focus on a set of business perspectives: Financial, Customer, Learning & Growth and Internal Business Processes. By defining objectives and measures for every perspective, the balanced scorecard helps to understand and pursue the overall strategic goal. The objectives are operationalized through a set of measures, targets and initiatives. As this paper deals with the aspect of delivering a technical BI solution, only measures and targets will be covered.

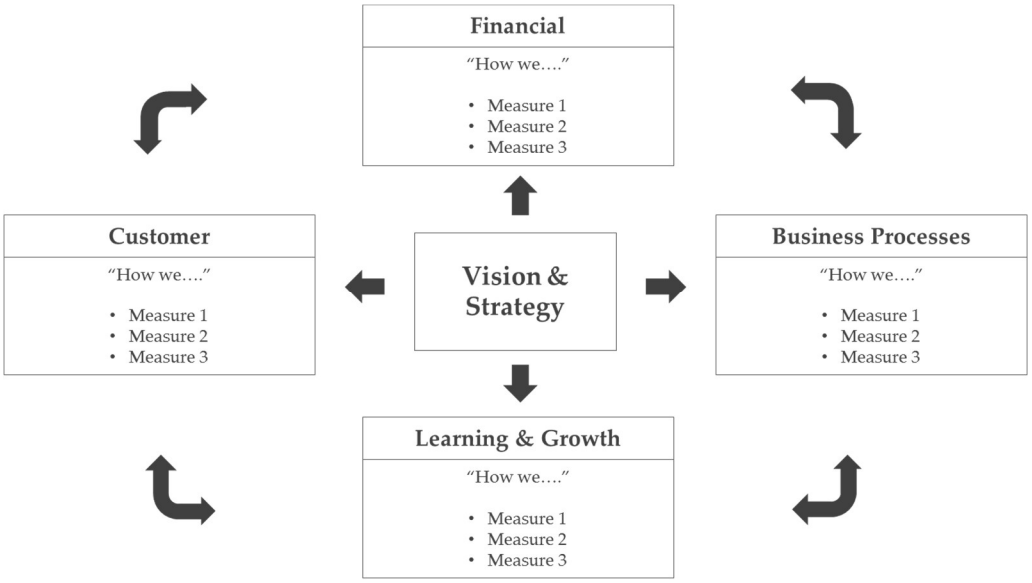


Figure 1: Balanced Scorecard

2.2 Balanced Scorecard for IMSports

For IMSports to continue its current track of growth, management has put focus on following measures:

Table 1: Balanced Scorecard for IMSports

Perspective	**Objective**	**Measures**
Financial	Satisfy shareholders with financial stability	Gross Margin, Monthly turnover
Customer	Increase customer retention and create ambassadors	Net Promoter Score (NPS), Complains Ratio, Customer Satisfaction
Internal Processes	Make sure products and staff help increasing efficiency	Employee Happiness, R&D Score, Breakage Rate
Learning & Growth	Maintain product innovation and competent staff	Product Innovation Rate, Product Margin, Staff Training

For explanation of measures, see appendix.

3 Data Warehouse

For the project at hand, the data warehouse made in the previous project was used. The data warehouse contains two fact tables, *FactSalesHeader* and *FactSalesDetails*, respectively describing the total figures and detailed product/quantity figures per transaction. Furthermore, there are six dimensions; *DimReason*, *DimLocation*, *DimTime*, *DimSalesPerson*, *DimCustomer* and *DimProduct*, describing sales reason, location of sales, time of sales, used sales channel, customer characteristics and product characteristics. It should be noticed, that products as a dimension of sales only can be measured with the fact table *FactSalesDetails*.

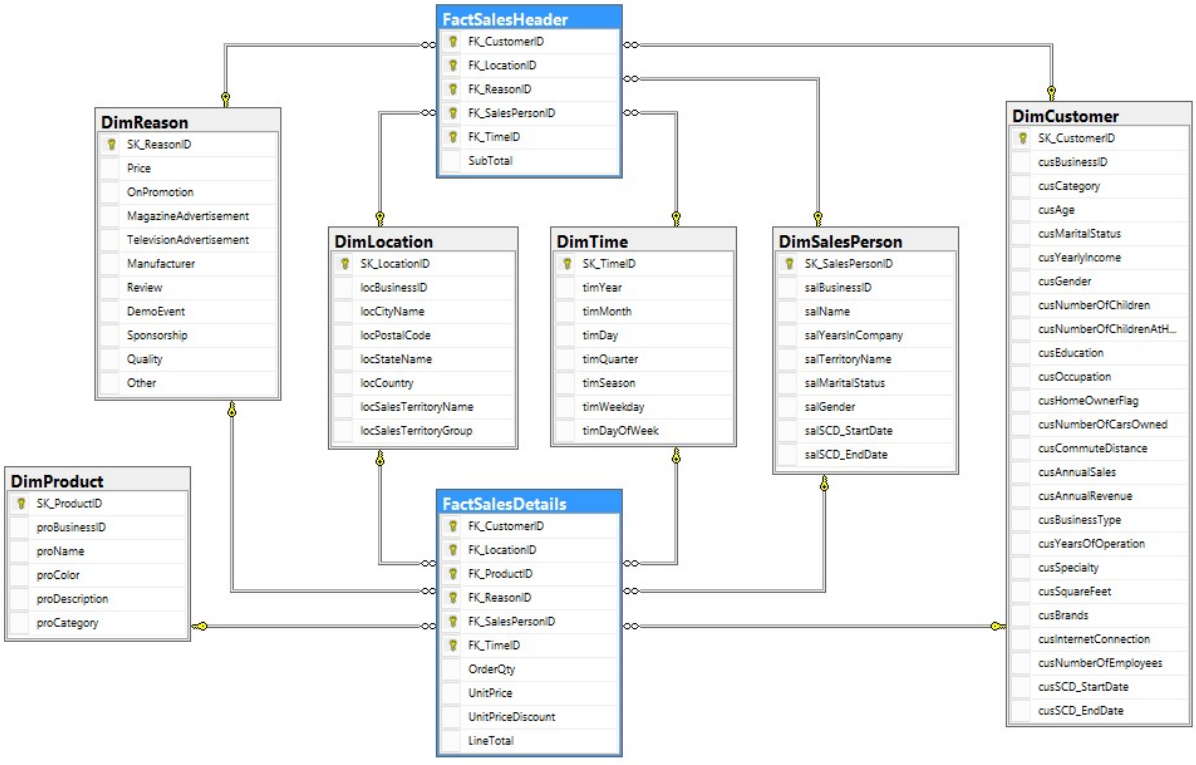


Figure 2: Used data warehouse

For the sake of adding content related to the business scenario, the following attributes were created artificially;

Table 2: Artificially Created Variables

Dimension	**Attribute**	**Description**
FactSalesHeader	SubCostTotal	The total cost of transaction line
	CustomerSatisfaction	1-10 score of customer satisfaction related to purchase
	Complain	Binary attribute expressing if customer has complained about purchase
FactSalesDetails	UnitCost	Cost per unit for the product given at the transaction line
	LineCostTotal	Order quantity multiplied with unit cost
DimProduct	proIsInnovative	Binary attribute expressing if product is considered an innovative product
	proReleaseYear	Release year of product
	proRDScore	1-10 score indicating how much research was spent on development of product
	proBreakageRate	Percentage of damaged products during production
DimSalesPerson	salStatus	Flag if employee is in training or not
	salTrainingExpenditure	Money spent training employee
	salEmployeeHappiness	1-10 score of employee happiness in latest survey
	salEmployeeExpertise	1-10 score of employee's expertise within sales

4 OLAP Cube

For creating a cube, the data warehouse was imported as a data source. The following data source view is presented:

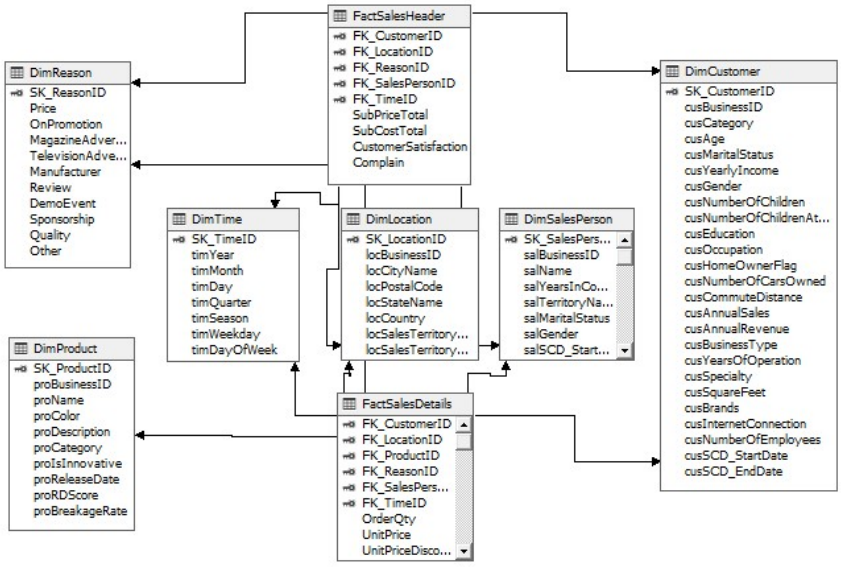


Figure 3: Data source view for OLAP Cube

4.1 Dimensions

For the dimensions to be suitable for analysis, each one was inspected and hierarchies were created.

For *DimTime* two hierarchies were created. First hierarchy was called *NumTimeHierarchy*, and contains numeric values for *timDay*, *timWeekday*, *timMonth*, *timQuarter* and *timYear*, where *timYear* is least granular and *timDay* most granular.

Second hierarchy was called *TextTimeHierarchy* and contains *timDayOfWeek* and *timSeason* describing respectively the name of each weekday and the name of each season.

Figure 4: Hierarchies for DimTime dimension

For *DimLocation* a hierarchy called *locHierarchy* was created, going from most granular level of *locCityName* to least granular level of *locSalesTerritoryGroup*, describing respectively name of the city and territory group (e.g. Pacific, Australia, Northern America)

Figure 5: Hierarchy for DimLocation

For *DimProduct* a hierarchy called *proHierarchy* was created, covering the product category as the least granular level and the product name.

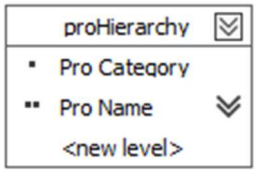


Figure 6: Hierarchy for DimProduct

For *DimSalesPerson* a hierarchy was built that takes into account the sales territory and the sales person's name.

Figure 7: Hierarchy for DimSalesPerson

4.2 Calculated Metrics

With created hierarchies, the *browse* function can be used to inspect measures in a cross-tab format. However, some measures are necessary to gain analytical insight. For that, the following calculated measures were created;

Table 3: Calculated Metrics

Fact Table	Calculated Member	Expression
Details	*LineProfitDetails*	`[Measures].[Line Price Total]-[Measures].[Line Cost Total]`
Header	*LineProfitHeader*	`[Measures].[Sub Price Total]-[Measures].[Sub Cost Total]`
	LineMarginHeader	`([Measures].[Sub Price Total]-[Measures].[Sub Cost Total])/[Measures].[Sub Price Total]`
	AvgCustomerSatisfaction Header	`Divide([Measures].[Customer Satisfaction], [Measures].[Fact Sales Header Count])`
	AvgComplainHeader	`Divide([Measures].[Complain], [Measures].[Fact Sales Header Count])`
	AvgLineProfitHeader	`Divide(LineProfit, [Measures].[Fact Sales Header Count])`
	AvgLineCostHeader	`Divide([Measures].[Sub Cost Total], [Measures].[Fact Sales Header Count])`

- *LineProfitDetails* express the profit by transaction line for *FactSalesDetails*
- *LineProfitHeader* express the profit by transaction line for *FactSalesHeader*
- *LineMarginHeader* express the margin in percent for *FactSalesHeader*
- *AvgCustomerSatisfaction* express the average customer satisfaction by transaction line for *FactSalesHeader*
- *AvgComplainHeader* express the transactions-to-complains ratio for *FactSalesHeader*
- *AvgLineProfitHeader* express the average profit by transactions line for *FactSalesHeader*
- *AvgLineCostHeader* express the average cost by transaction line for *FactSalesHeader*

4.3 Named Sets

For management to distinguish between the two sales channels, internet sales and human sales, when using *browse* function, a named set called *InternetSalesSet* was created with purpose of selecting it (showing internet sales) and deselecting it (showing human sales). It was built with the following expression:

```
Exists(
[Dim Sales Person].[Sal Name].[Sal Name].Members,
[Dim Sales Person].[Sal Name].[Sal Name].&[**Direct Purchase**])
```

4.4 Key Performance Indicators

For IMSports to follow up on the most important measures, financials and customers, three KPI's were created: *CustomerSatisfactionKPI, SatisfiedCustomersKPI* and *MarginKPI.* Management has pointed out, that the KPI's need to be differentiated based on sales channel.

CustomerSatisfactionKPI was created on top of the previous created calculated metric *AvgCustomerSatisfaction*. Management argues that average customer satisfaction goal for internet sales is 7, while being 9 for human sales and 8 in total, considering that a human interaction expectably is better than solely computer interaction when ordering. The following goal expression was used:

```
CASE WHEN [Dim Sales Person].[sal Name].CurrentMember IS
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 7
WHEN [Dim Sales Person].[sal Name].CurrentMember <>
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 9
ELSE 8
END
```

With status expression:

```
CASE
WHEN KPIVALUE("CustomerSatisfactionKPI")/KPIGOAL("CustomerSatisfactionKPI") >1
THEN 1
WHEN KPIVALUE("CustomerSatisfactionKPI")/KPIGOAL("CustomerSatisfactionKPI") <1
```

```
AND KPIVALUE("CustomerSatisfactionKPI")/KPIGOAL("CustomerSatisfactionKPI") >0.8
THEN 0
ELSE -1
END
```

SatisfiedCustomerKPI was built with the expression 1-[Measures].[AvgComplainHeader], reflecting the number of customers in average that haven't complained.

Management argues that the goal of non-complaining customers is 70% for internet sales, while being 90% for human sales and 80% in total. The following goal expression is used:

```
CASE WHEN [Dim Sales Person].[sal Name].CurrentMember IS
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 0.70
WHEN [Dim Sales Person].[sal Name].CurrentMember <>
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 0.80
ELSE 0.75
END
```

With status expression:

```
CASE
WHEN KPIVALUE("SatisfiedCustomersKPI")/KPIGOAL("SatisfiedCustomersKPI") >=1
THEN 1
WHEN KPIVALUE("SatisfiedCustomersKPI")/KPIGOAL("SatisfiedCustomersKPI") >0.80
AND KPIVALUE("SatisfiedCustomersKPI")/KPIGOAL("SatisfiedCustomersKPI") <1
THEN 0
ELSE -1
END
```

MarginKPI was created on top of the previous created calculated metric *LineProfitHeader*. Management argues that average margin for internet sales is 50%, while being 30% for human sales and 40% in total, considering that internet sales channel distributes more profitable products while products related to human sales are costlier. The following goal expression was used:

```
CASE WHEN [Dim Sales Person].[sal Name].CurrentMember IS
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 0.5
WHEN [Dim Sales Person].[sal Name].CurrentMember <>
[Dim Sales Person].[sal Name].&[**Direct Purchase**]
THEN 0.3
ELSE 0.4
```

```
END
```

With status expression:

```
CASE
WHEN KPIVALUE("MarginKPI")/KPIGOAL("MarginKPI") >=1
THEN 1
WHEN KPIVALUE("MarginKPI")/KPIGOAL("MarginKPI") >0.80
AND KPIVALUE("MarginKPI")/KPIGOAL("MarginKPI") <1
THEN 0
ELSE -1
END
```

Inspecting the KPI's for internet sales, it is overperforming on all three parameters.

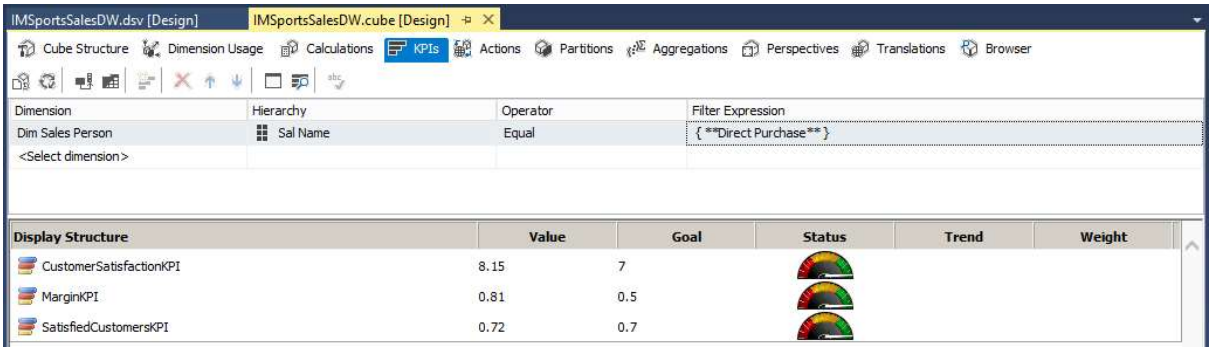


Figure 8: KPI's for internet sales

Inspecting the status KPI's for human sales, it is overperforming on margin but lack performance on customer satisfaction and customer complains.

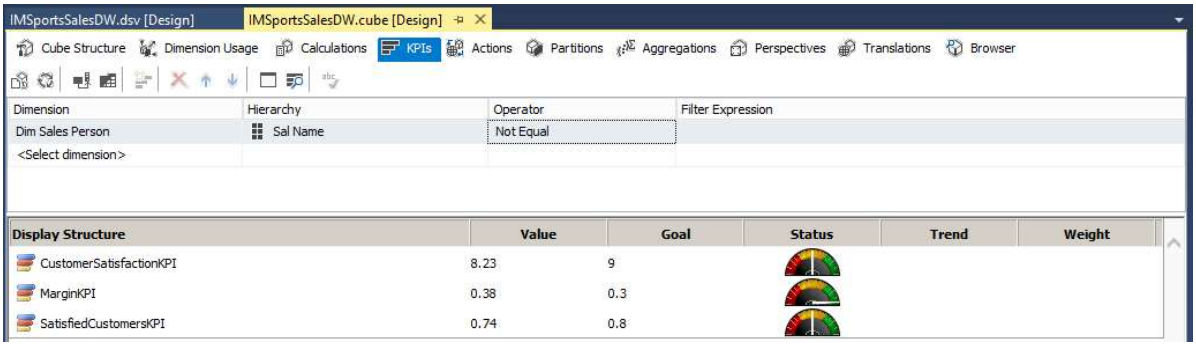


Figure 9: KPI's for human sales

4.5 Partitions

The purpose of partitions is to split data in smaller quantities making it manageable and improving query performance. Partitioning must be logically structured, as overlapping data causes redundancy and wastes storage. Thus, *DimDate* is the dimension used for partitioning data, defining a clear border for data splits, with the attribute *timYear*.

FactSalesHeader contains 31.422 rows by default, making it heavy to operate, and thus it was decided to partition data into four chunks based on *timYear*. The following query was used for partitioning when choosing 'query binding':

```
SELECT
[dbo].[FactSalesHeader].[FK_CustomerID],[dbo].[FactSalesHeader].[FK_LocationID],[dbo].[
FactSalesHeader].[FK_ReasonID],[dbo].[FactSalesHeader].[FK_SalesPersonID],[dbo].[FactSa
lesHeader].[FK_TimeID],[dbo].[FactSalesHeader].[SubPriceTotal],[dbo].[FactSalesHeader].
[SubCostTotal],[dbo].[FactSalesHeader].[CustomerSatisfaction],[dbo].[FactSalesHeader].[
Complain]
FROM [dbo].[FactSalesHeader]
INNER JOIN DimTime ON DimTime.SK_TimeID = FactSalesHeader.FK_TimeID
WHERE DimTime.timYear = [GIVEN YEAR]
```

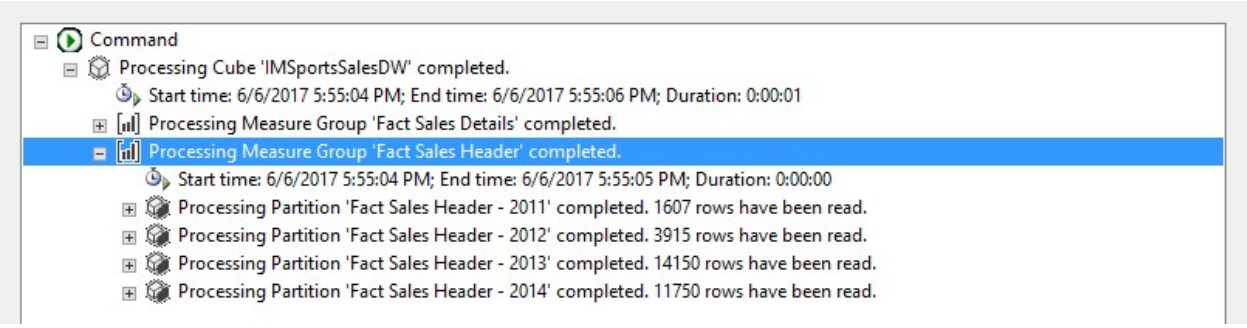
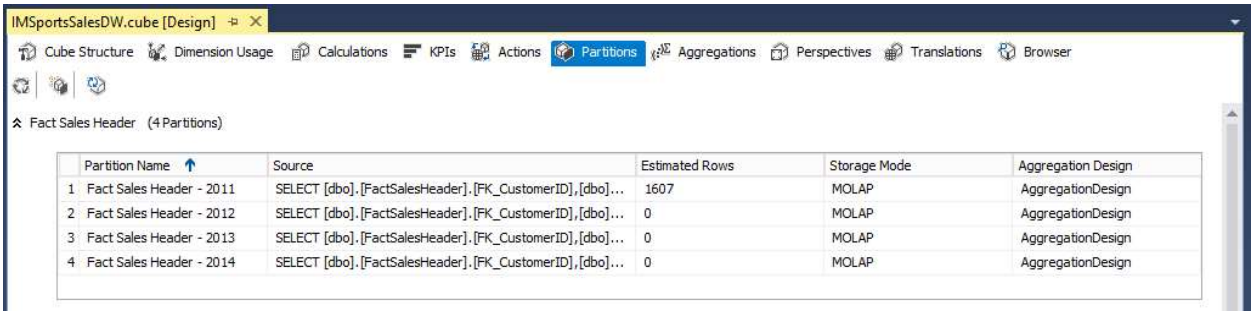


Figure 10: Partitioning being processed

For the year 2011, 1607 rows were partitioned in to a separate part, while it was 3915, 14150 and 11750 rows, making it respectively 12.5%, 45% and 37.5% of the data for the years 2012, 2013 and 2014.



	Partition Name	Source	Estimated Rows	Storage Mode	Aggregation Design
1	Fact Sales Header - 2011	SELECT [dbo].[FactSalesHeader].[FK_CustomerID],[dbo]...	1607	MOLAP	AggregationDesign
2	Fact Sales Header - 2012	SELECT [dbo].[FactSalesHeader].[FK_CustomerID],[dbo]...	0	MOLAP	AggregationDesign
3	Fact Sales Header - 2013	SELECT [dbo].[FactSalesHeader].[FK_CustomerID],[dbo]...	0	MOLAP	AggregationDesign
4	Fact Sales Header - 2014	SELECT [dbo].[FactSalesHeader].[FK_CustomerID],[dbo]...	0	MOLAP	AggregationDesign

Figure 11: Partitioning of FactSalesHeader

4.6 Perspectives

Again, although not improving cube performance, a subset of data perspectives can be created for the purpose of focusing on a business-specific viewpoint. As the cross-tab *browse* function can be complex for users, a total of three perspectives was created: *Customer and Location Perspective, Product Performance Perspective* and *Sales Person Performance Perspective.*

The *Customer and Location Perspective* deals with measuring how sales are distributed by location of the customers and for what reason customers are buying. For that, *DimReason* and *DimCustomer* are activated as available dimensions while using *FactSalesHeader* as fact table with related calculated members.

The *Product Performance Perspective* deals with measuring order quantities, product margin and distribution of product sales over time. For that, *DimTime* and *DimProduct* are activated as available dimensions, while using *FactSalesDetails* as fact table and with related calculated member *LineProfitDetails*.

The *Sales Person Performance Perspective* deals with measuring how sales are distributed by sales person over time. For that, *DimSalesPerson* and *DimTime* are activated as available dimensions, while using *FactSalesHeader* as fact table. All calculated members related to *FactSalesHeader* is furthermore activated.

Cube Objects	Object Type	Perspective Name	Perspective Name	Perspective Name
IMSportsSalesDW	Name	Product Performance Per…	Sales Person Performanc…	Customer and Location P…
	DefaultMeasure			
− Measure Groups				
− Fact Sales Header	MeasureGroup	☐	☑	☑
Sub Price Total	Measure	☐	☑	☑
Sub Cost Total	Measure	☐	☑	☑
Customer Satisfaction	Measure	☐	☑	☑
Complain	Measure	☐	☑	☑
Fact Sales Header Count	Measure	☐	☑	☑
− Fact Sales Details	MeasureGroup	☑	☐	☐
Order Qty	Measure	☑	☐	☐
Unit Price	Measure	☑	☐	☐
Unit Price Discount	Measure	☑	☐	☐
Line Price Total	Measure	☑	☐	☐
Unit Cost	Measure	☑	☐	☐
Line Cost Total	Measure	☑	☐	☐
Fact Sales Details Count	Measure	☑	☐	☐
− Dimensions				
+ Dim Time	CubeDimension	☑	☑	☐
+ Dim Sales Person	CubeDimension	☐	☑	☐
+ Dim Location	CubeDimension	☐	☐	☑
+ Dim Customer	CubeDimension	☐	☐	☑
+ Dim Reason	CubeDimension	☐	☐	☐
+ Dim Product	CubeDimension	☑	☐	☐
− KPIs				
CustomerSatisfactionKPI	Kpi	☐	☑	☑
SatisfiedCustomersKPI	Kpi	☐	☑	☑
MarginKPI	Kpi	☐	☑	☑
− Calculations				
InternetSalesSet	NamedSet	☐	☑	☑
LineProfitHeader	CalculatedMem…	☐	☑	☑
LineMarginHeader	CalculatedMem…	☐	☑	☑
LineProfitDetails	CalculatedMem…	☑	☐	☐
AvgCustomerSatisfactionHeader	CalculatedMem…	☐	☑	☑
AvgComplainHeader	CalculatedMem…	☐	☑	☑
AvgLineProfitHeader	CalculatedMem…	☐	☑	☑
AvgLineCostHeader	CalculatedMem…	☐	☑	☑

Figure 12: Perspective settings

4.7 Data Mining Capabilities

Finishing the analytical part of SSAS, the management of IMSports is interested in clustering the private customers for them to understand the characteristics of the customers. A cluster model is used in SSAS data mining capabilities.

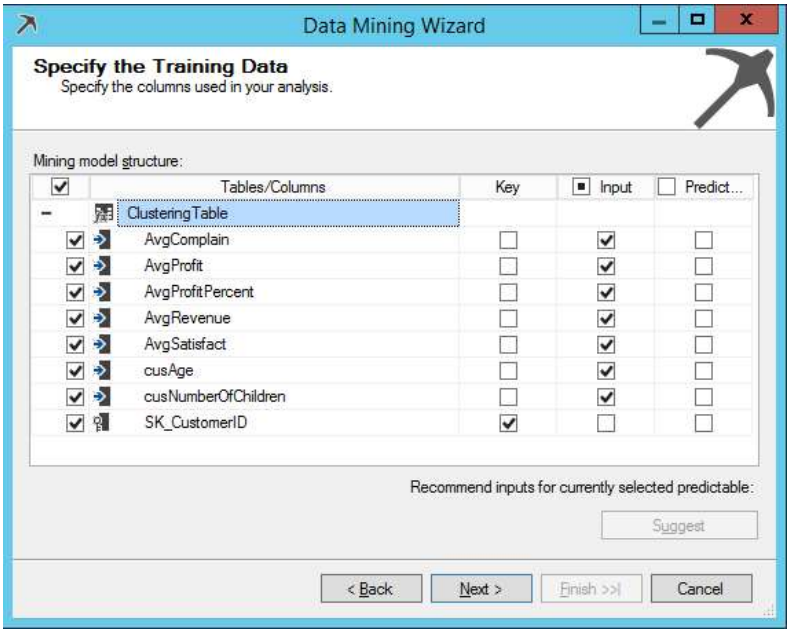


Figure 13: Selecting input variables for clustering

Seven characteristic variables were chosen as input variables.

The outcome is 10 clusters that are shown in the cluster diagram with density visualized by color saturation. The color saturation of the link between the clusters, reflects the distance between the clusters. Furthermore, cluster characteristics are available amongst many other features, such as cluster profiling (distributions of variables) and discrimination of variables.

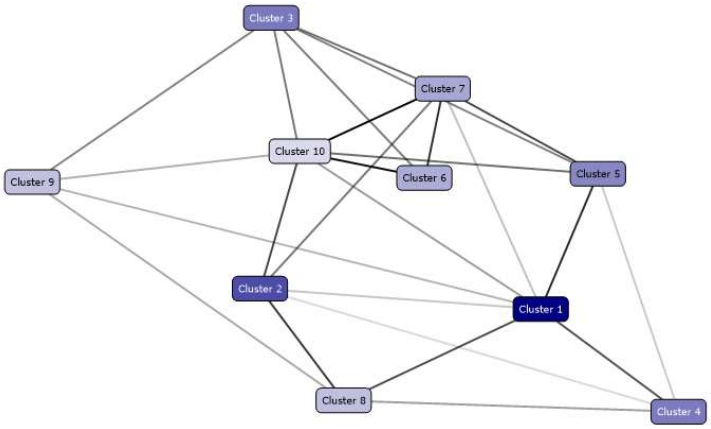


Figure 14: Clusters and cluster links visualized

Customers going into cluster 1 have a high probability of low profit, low revenue, high customer satisfaction and low amount of complains.

Characteristics for Cluster 1

Variables	Values	Probability
Avg Profit	-14.6 - 103.9	
Avg Revenue	2.3 - 207.4	
Avg Complain	0.0	
Avg Profit Percent	0.2 - 0.5	
Avg Profit Percent	0.5 - 0.7	
Cus Number Of Children	0	
Avg Satisfact	8.2 - 9.2	

Figure 15: Cluster characteristics for cluster 1

Customers going into cluster 3 have a high probability of high profit, high margin, high customer satisfaction and low amount of complains.

Characteristics for Cluster 3

Variables	Values	Probability
Avg Profit	721.0 - 1,338.1	
Avg Profit Percent	0.5 - 0.7	
Avg Revenue	907.8 - 1,608.2	
Avg Complain	0.0	
Avg Revenue	1,608.2 - 3,578.3	
Avg Profit Percent	0.7 - 0.9	
Avg Satisfact	8.2 - 9.2	
Cus Number Of Children	0	

Figure 16: Cluster characteristics for cluster 3

5 Reporting

Reporting is the process of collecting and delivering insightful data, requested by management, reflecting activities in a business. Reporting is in this case supported by SQL Server Reporting Services (SSRS), Excel and Power BI, making it feasible to create everything from simple reports to more advanced dashboards.

5.1 Reporting in SSRS

Top management of IMSports has, on behalf of middle management, expressed requests for a set of simple reports to be delivered to middle management, making it possible for the department managers to follow performance of the employees and state of the department. Thus, department managers from financials, sales and product development have requested reports.

First report, *Complains by Country and Year*, is made with the purpose of displaying the development in complains, relative to the number of orders, by country. A filter function makes it possible for the user to interactively compare complains ratio for countries.

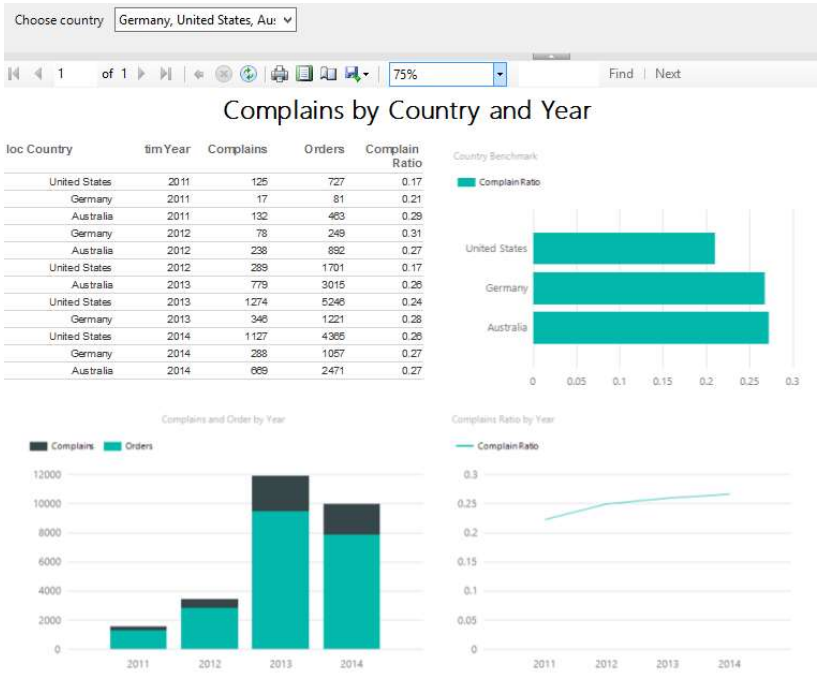


Figure 17: Complains by Country and Year in SSRS

Second report, *Complains by Sales Person and Year*, is made with the same purpose as stated in previous example, just grouped by sales staff. Two filter functions make it possible to select both year and sales person, so that user can triangulate the report from a customized perspective. Moreover, a sort function is made, enabling users to sort by name, year or complains ratio.

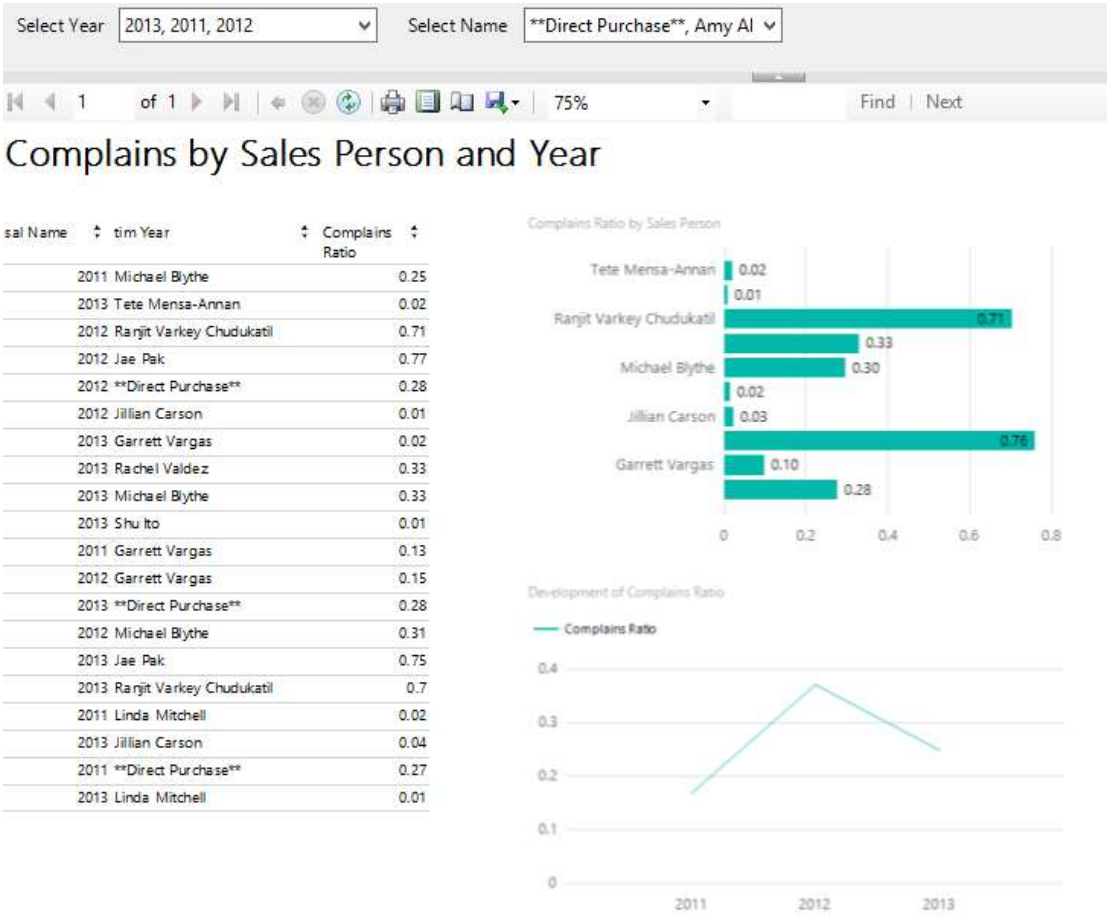


Figure 18: Complains by Sales Person and Year in SSRS

Third report, *Financials by Year and Month*, is made with the purpose of showing the development in turnover, cost and gross margin, by year and month. Furthermore, a trend chart is available for each year, indicating the overall trend for the given year. For instance, it is easy to understand the development of gross margin for year 2013, as the trend chart show an increasing tendency.

Financials by Year and Month

Turnover

Year	1	2	3	4	5	6	7	8	9	10	11	12	Trend	Total
2011					504K	459K	2045K	2496K	502K	4589K	738K	1310K		12641K
2012	3971K	1475K	2976K	1635K	3075K	4099K	3418K	2176K	3454K	2544K	1873K	2829K		33524K
2013	2088K	2317K	3412K	2532K	3246K	5081K	4896K	3334K	4533K	4796K	3312K	4075K		43621K
2014	4290K	1338K	7217K	1797K	5366K	49K								20057K
Total	10348K	5130K	13605K	5964K	12190K	9688K	10359K	8005K	8489K	11928K	5922K	8215K		109844K

Cost

Year	1	2	3	4	5	6	7	8	9	10	11	12	Trend	Total
2011					266K	38K	678K	884K	42K	1752K	63K	314K		4037K
2012	1117K	464K	1020K	770K	2828K	4527K	3534K	1893K	3040K	2280K	1084K	1621K		24178K
2013	1324K	1951K	2843K	2333K	2559K	3959K	4037K	1848K	2716K	2844K	1417K	1984K		29816K
2014	2046K	348K	4421K	440K	2993K	33K								10280K
Total	4488K	2762K	8283K	3542K	8647K	8557K	8249K	4625K	5799K	6876K	2564K	3919K		68311K

Gross Margin

Year	1	2	3	4	5	6	7	8	9	10	11	12	Trend	Total
2011					237K	421K	1366K	1612K	460K	2837K	675K	996K		8604K
2012	2853K	1012K	1956K	865K	247K	-428K	-116K	283K	414K	264K	788K	1209K		9346K
2013	763K	366K	570K	199K	687K	1122K	859K	1485K	1817K	1952K	1894K	2091K		13805K
2014	2244K	990K	2796K	1357K	2373K	16K								9777K
Total	5860K	2368K	5322K	2421K	3544K	1132K	2109K	3380K	2690K	5052K	3358K	4295K		41532K

Figure 19: Financials by Year and Month in SSRS

As previous reports in SSAS only displayed the customer satisfaction as a single measure, the Net Promoter Score (NPS) is now calculated, with the formula:

$$Net\ Promoter\ Score = \frac{Promoters - Detractors}{Total\ responses}$$

Where:

$$Promoters = customers\ giving\ 9{:}10\ in\ customer\ satisfaction\ score$$
$$Detractors = customers\ giving\ 1{:}6\ in\ customer\ satisfaction\ score$$
$$Total\ responses = number\ of\ transactions\ with\ a\ customer\ satisfaction\ score$$

The second set of reports, *NPS by Salesman and Year* and *NPS by Country and Year,* focuses on development of NPS and orders by sales person and country.

NPS by Salesman and Year allows the user to expand and collapse the different years in a drilldown, making the report more flexible. The internet sales (**Direct Purchase**) has a very poor NPS score, but has a great share of orders for the year of 2011, whereas the human sales channel perform significant better for NPS but a lower share of orders.

NPS by Salesman and Year

Year	Sales Person	NPS	Orders
⊟ 2011			
	Direct Purchase	0.31	1201
	David Campbell	1	28
	Garrett Vargas	0.83	30
	Jillian Carson	1	59
	José Saraiva	1	56
	Linda Mitchell	0.98	46
	Michael Blythe	0.38	65
	Pamela Ansman-Wolfe	1	22
	Shu Ito	1	33
	Stephen Jiang	1	4
	Tsvi Reiter	1	63
⊞ 2012			
⊞ 2013			
⊞ 2014			

Figure 20: NPS report by Salesman and Year

Last report, *NPS by Country and Year*, focus on development of NPS and orders throughout time and country. The two charts to the left are focused on number of orders, whereas the two charts to the right are focused on NPS. The charts clearly show that United States has the highest share of orders, and they furthermore have the greatest NPS score.

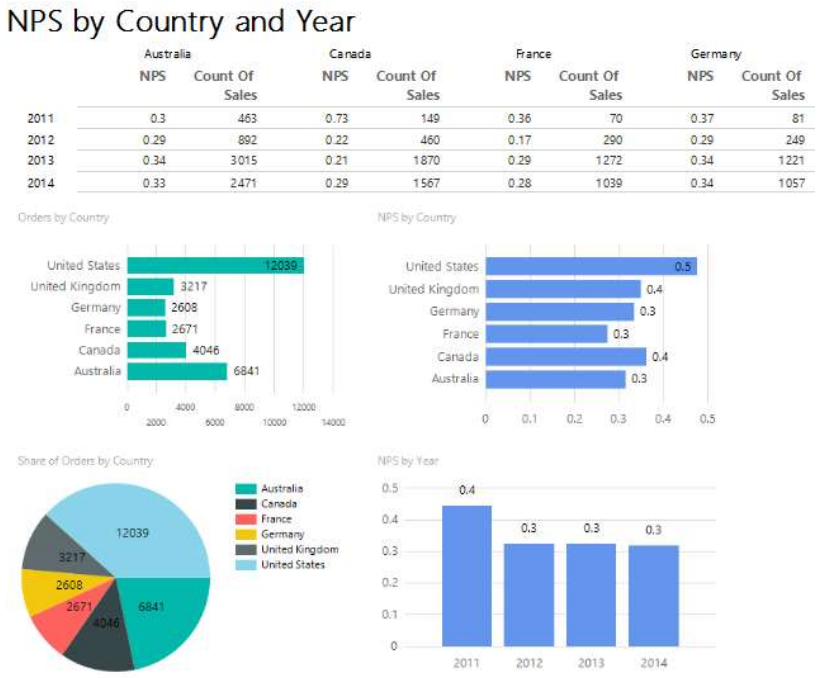


Figure 21: NPS report by Country and Year

Business Conclusion:

There are four sales people – namely Pak, Blythe, Chudukatil and Ito – that generate most overall complaints. Also, the number of complaints over the internet channel skyrocketed in 2013 and has not recovered ever since. Consequently, the complaints rate in any country increases over time, with the US being a minor exception, since the relative number of internet purchases is lower. The same accounts for the initially excellent net promoter score that drops significantly over time. Overall, Pak and Chudukatil have exceptionally low NPS. It might be advisable to remove them from further operations and improve the overall perceived quality of the internet channel.

5.2 Reporting in Excel – Simple Pivot Tables

For simple reporting, like in SSRS, Excel offers pivot tables capable of connecting directly to a cube. Furthermore, the previously built perspectives are available, limiting users to work with narrower perspectives of the cube. In this case, we selected the original cube.

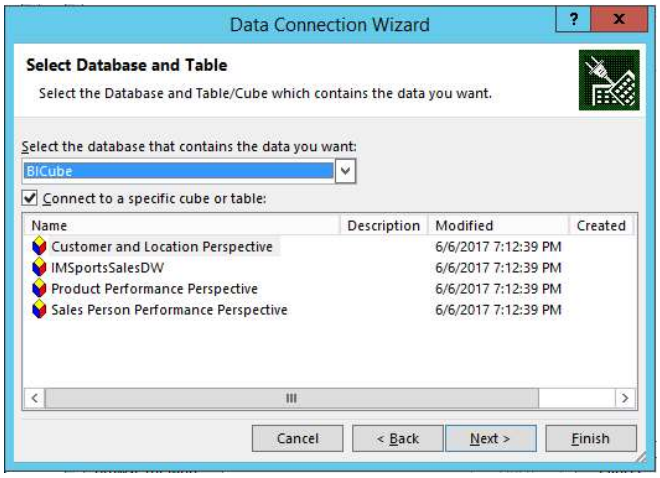


Figure 22: Selecting data source for pivot table

With a connection established, a report inspecting the turnover, cost, order quantities and average unit cost was created, with a filter for innovative products (*proIsInnovative*).

proIsInnovativ True

Row Labels	Sum of LinePriceTotal	Sum of LineCostTotal	Sum of OrderQty	Average of UnitCost
PROD: Spiky M	1,721,243	864,404	1,120	874
PROD: Fast & F	1,657,198	831,759	1,477	656
PROD: Sleepy I	1,426,373	738,513	1,005	863
PROD: Smokin	1,120,066	249,924	944	265
PROD: Wheeli	981,188	504,231	652	890
PROD: Sympat	635,724	318,108	820	421
PROD: Hill Clin	438,867	219,511	538	448
PROD: Massive	358,122	185,049	851	238
PROD: Aggresi	290,075	149,973	686	240
PROD: Furious	241,774	120,887	425	321
PROD: Yelling I	161,293	82,658	515	174
PROD: Ghostri	157,569	78,785	449	188
PROD: Slowly T	145,089	74,014	450	177
PROD: King of	122,512	62,544	381	176
PROD: Flying D	101,734	50,867	282	197
PROD: Slowly A	96,982	63,308	272	233
PROD: Speedy	44,484	22,269	813	27
Grand Total	**9,700,295**	**4,616,804**	**11,680**	**443**

Figure 23: Pivot table report for Innovative Products

Furthermore, a chart was created to display the turnover of the products, ordering turnover descending.

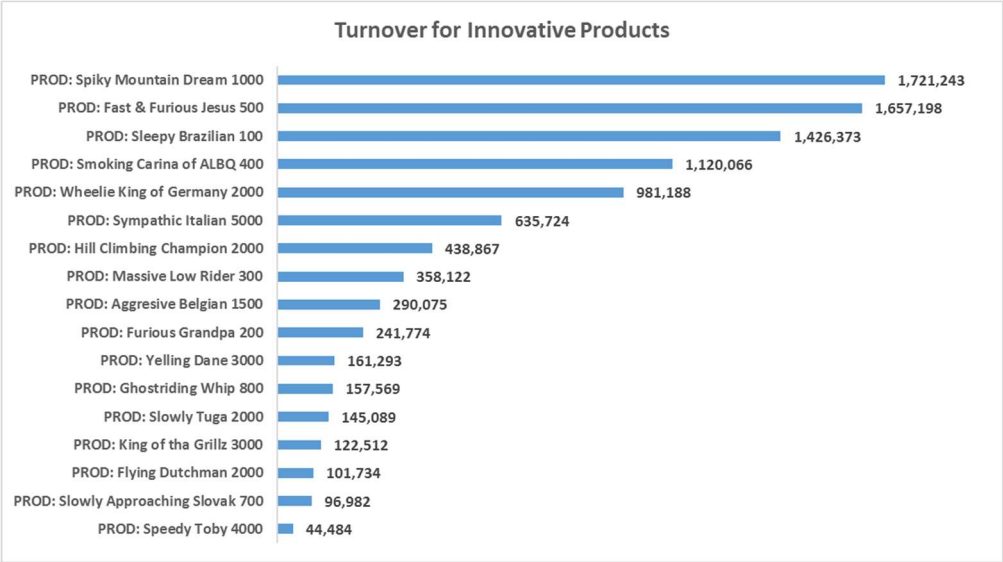


Figure 24: Turnover for Innovative Products

However, simple pivot tables in Excel have a limitation, as they cannot create new calculations in the pivot table, unless the specific measure is built as an attribute in the source (Cube). In this case, management would like a new bar per product, displaying the profit per product. For that purpose, Power Pivot is useful.

Business Conclusion:

There are four innovative products that have evolved to crucial items, showing high sales quantities and high sales prices and revenues over 1 million €.

5.3 Reporting in Excel – Power Pivot

Like classic pivot tables, Power Pivot in Excel allows the user to connect directly to an external data source. However, the user has the possibility to manipulate and construct tables, without editing the data source. For demonstration purposes, a table was created to the right (*FactSalesDetailsExtra)*, as shown in the diagram.

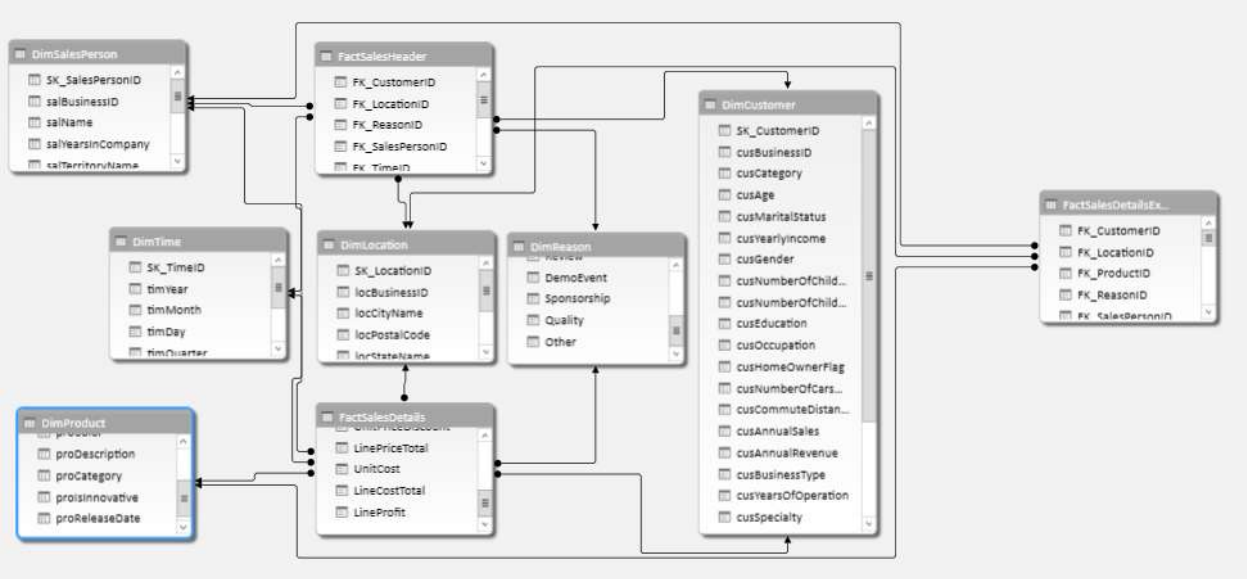


Figure 25: Data view in Power Pivot

Thus, in *data view* in Power Pivot, the *FactSalesDetailsExtra* was copied from *FactSalesDetails*, and had some attributes from the dimension tables added and some simple calculations created. These can as well be constructed in the actual fact table. Amongst them were:

Attribute	**Formula**
salCategory	=IF([salName]="**Direct Purchase**","Internet","Human")
LineProfit	=[LinePriceTotal]-[LineCostTotal]
proIsInnovative	=RELATED(DimProduct[proIsInnovative])

With *LineProfit* created, as a measure of profit per line in EUR, it is possible to display the profit per product in a report. However, for measures to be added to a Power Pivot table,

it is not required to create a new attribute. Like in SSAS, a calculated field (member) can be created. To demonstrate the profit in percent, a calculated field was created, called *MarginPercent*.

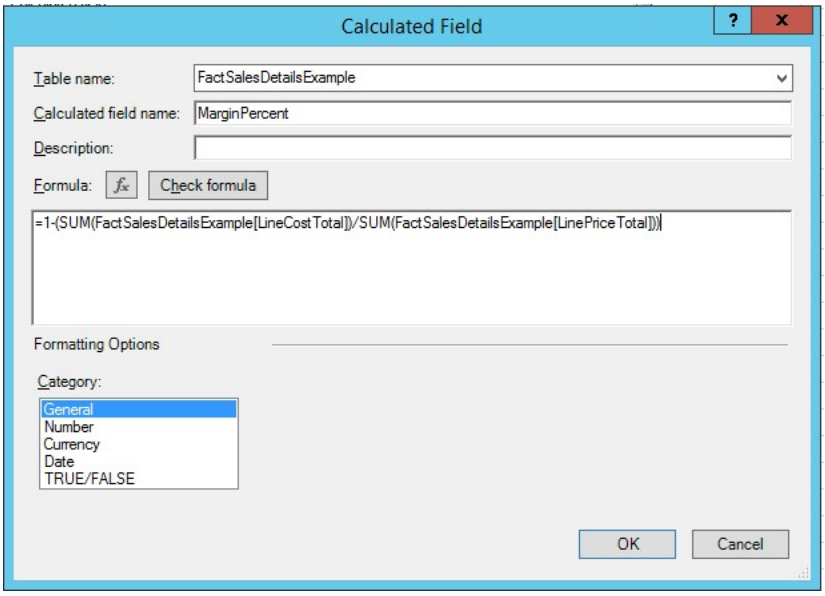


Figure 26: Creating calculated field in Power Pivot

With two extra fields, *LineProfit* and *MarginPercent*, a new attribute and calculated measure respectively, Power Pivot displays a pivot table, that standard Excel isn't capable of.

prolsInnovative	True

Row Labels	Sum of LinePriceTotal	Sum of LineCostTotal	Sum of OrderQty	Average of UnitCost	Sum of LineProfit	MarginPercent
PROD: Spiky Mountai	1,721,243	864,404	1120	$873.85	856,838	0.50
PROD: Fast & Furious	1,657,198	831,759	1477	$655.71	825,440	0.50
PROD: Sleepy Brazilia	1,426,373	738,513	1005	$863.30	687,860	0.48
PROD: Smoking Carin	1,120,066	249,924	944	$264.75	870,142	0.78
PROD: Wheelie King c	981,188	504,231	652	$890.19	476,957	0.49
PROD: Italian Fireball	635,724	318,108	820	$421.23	317,615	0.50
PROD: Hill Climbing C	438,867	219,511	538	$447.69	219,356	0.50
PROD: Massive Low Ri	358,122	185,049	851	$238.28	173,073	0.48
PROD: Aggresive Belg	290,075	149,973	686	$240.27	140,102	0.48
PROD: Furious Grandp	241,774	120,887	425	$320.87	120,887	0.50
PROD: Yelling Dane 3(	161,293	82,658	515	$174.29	78,635	0.49
PROD: Ghostriding Wl	157,569	78,785	449	$188.07	78,785	0.50
PROD: Slowly Tuga 20	145,089	74,014	450	$176.66	71,076	0.49
PROD: King of tha Gril	122,512	62,544	381	$176.45	59,968	0.49
PROD: Flying Dutchm	101,734	50,867	282	$196.56	50,867	0.50
PROD: Slowly Approa	96,982	63,308	272	$232.75	33,674	0.35
PROD: Speedy Toby 4(	44,484	22,269	813	$27.43	22,216	0.50
Grand Total	**9,700,295**	**4,616,804**	**11680**	**$443.15**	**5,083,490**	**0.52**

Figure 27: Pivot Table report made in Power Pivot

A report and chart was created for the product development manager, displaying the turnover and the profit for each innovative product. Thus, the product development manager can tell that sales does not equal profitability.

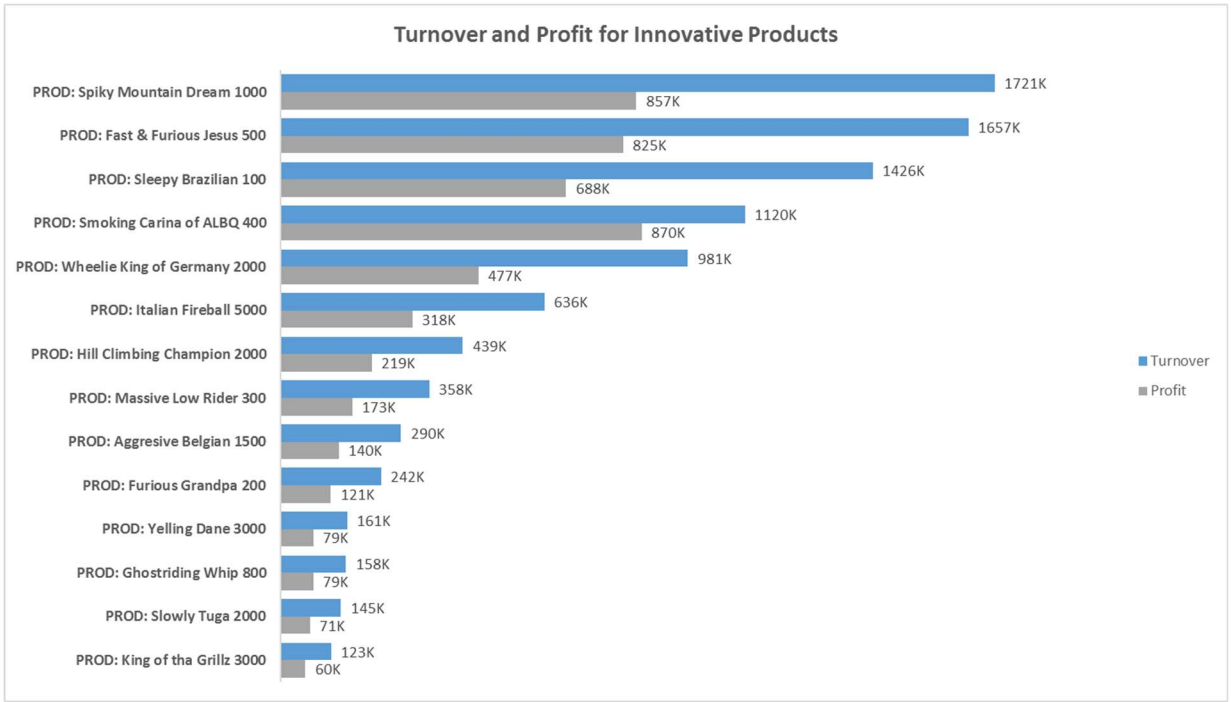


Figure 28: Turnover and Profit for Innovative Products

Furthermore, a report showing the calculated field *MarginPercent* for innovative (true) and non-innovative products (false) by the years 2013 and 2014, was created.

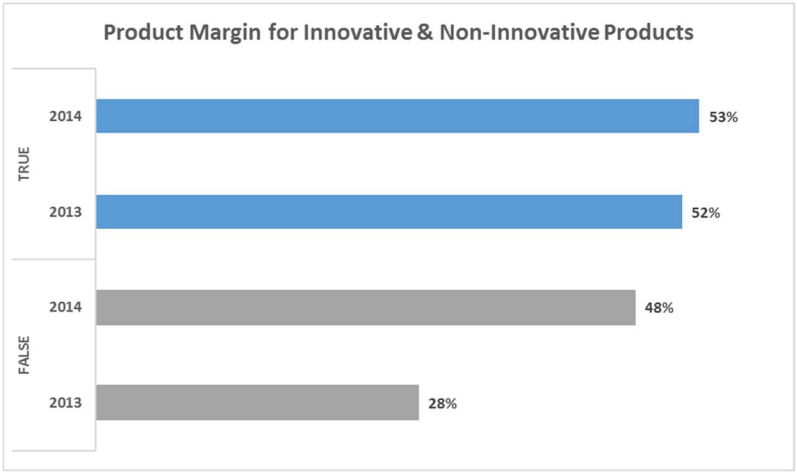


Figure 29: Product Margin for Innovative and Non-Innovative Products

Business Conclusion:

Although showing excellent revenue, the previously highlighted items do not show the highest profitability, which could be caused by high research, development and production cost. One outstanding product is SCA 400 with a gross production profitability of 77.7%. Also, innovative products show a higher product margin than conservative ones. The margin for conservative products is highly variant as well, showing the competitiveness for such commodity products on the market.

5.4 Reporting in Power BI

For detailed reporting to top management, allowing them to follow reporting for the entire business, while diving into specific departments and measures, Power BI is useful.

Five pages of dashboards were built, with the first being a scorecard based on the balanced scorecard philosophy, while the remaining four pages are focused on each of the categories of the balanced scorecard. Thus, the first page, called *Scorecard*, gives management the option to follow each aspect of the balanced scorecard, while adjusting the filters on the left side. In total 15 measures were constructed, that enable top management to follow targets for the balanced scorecard.

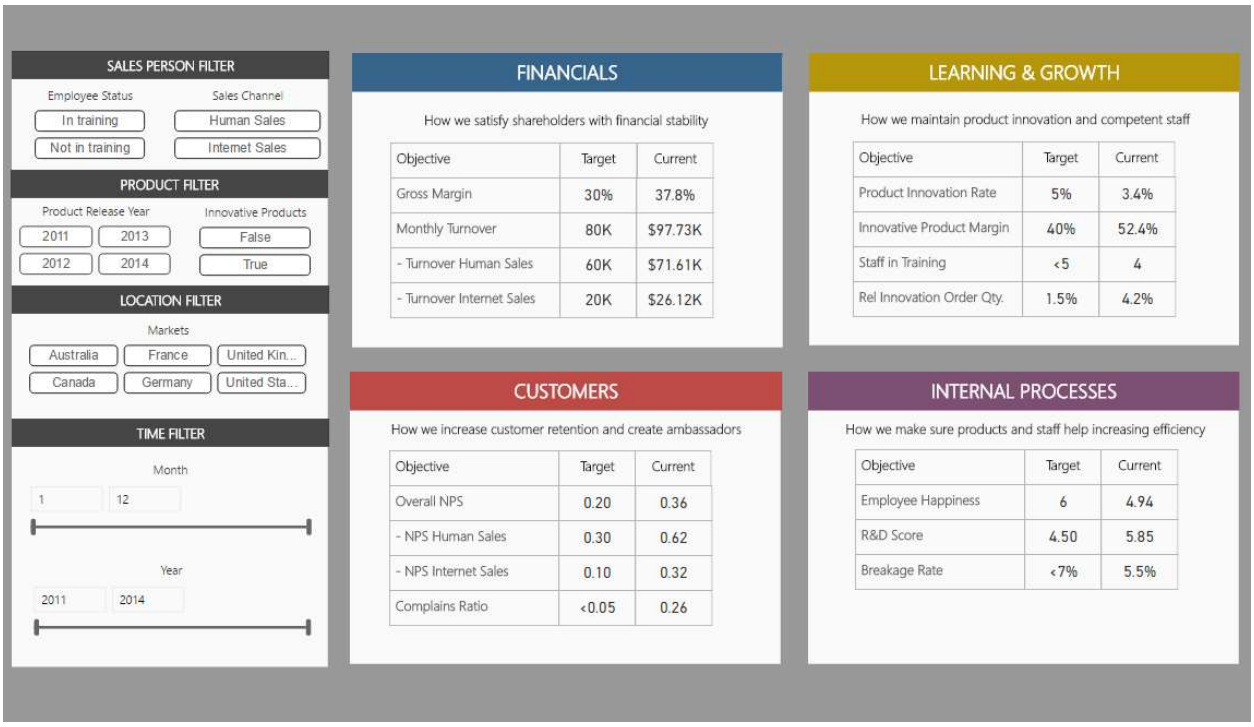


Figure 30: Balanced Scorecard presented in Power BI

Business Conclusion:

On the first look, the management can verify that IMSports is financially well established, not only reaching but confidently exceeding the targeted values. Customers seem relatively satisfied with the provided service, however there is a concerning relative

number of complaints. Internally, the company shows a healthy structure with a low breakage rate but the employee happiness needs to be improved. Finally with regards to learning and growth, the target for product innovation rate was not reached.

Financial Perspective:

The second page, *Financials*, tracks financial goals. The left side of the page offers insight in the primary financial figures, such as turnover, cost, profit, average order quantities and average unit price.

The right side offers a gauge measuring margin as turnover subtracted with cost. Two filters are furthermore available, for selecting a certain time span of years and/or months.

Below the gauge, the financial measures are split in to three categories; markets, products and sales persons, giving insight on sales on a more specific level.

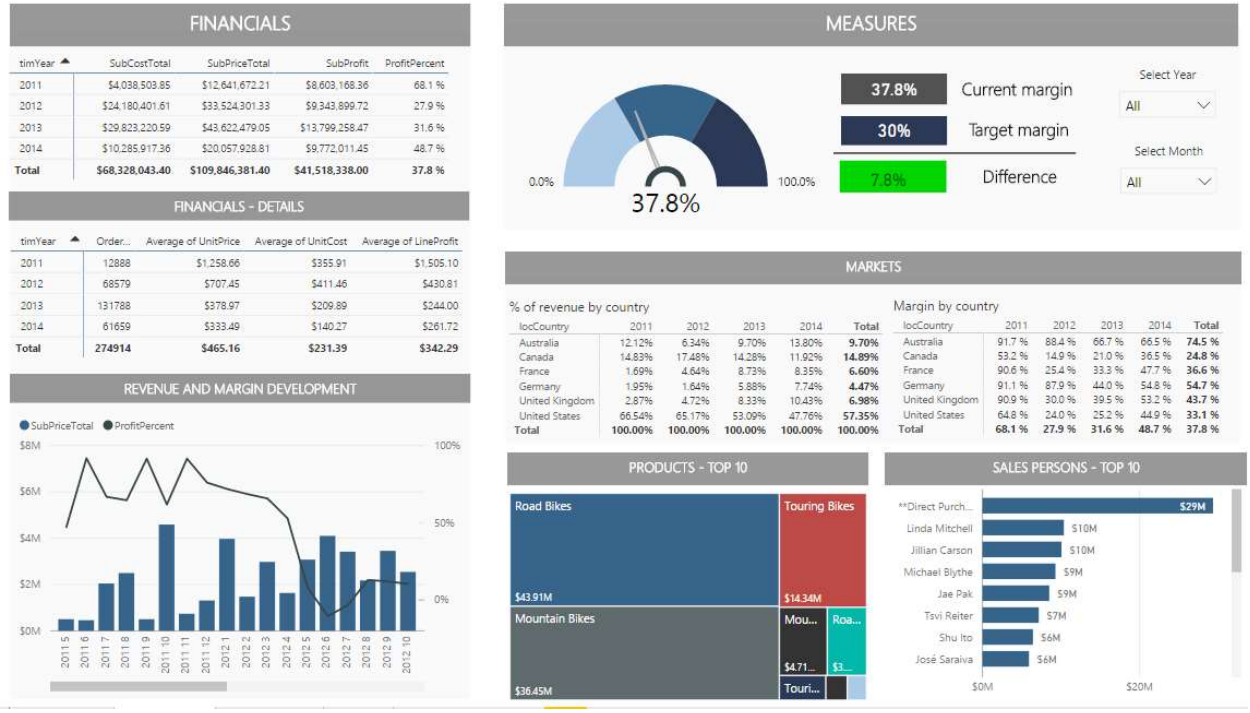


Figure 31: Financial Scorecard in Power BI

Business Conclusion:

As seen on the BSC cockpit, IMSports performs well on the financial perspective. The company shows a constant growth in revenue and profit over the last years and is currently performing on a margin that is better than targeted for. A nice development is

the outreaching globalization of sales, with a relatively decreasing number of domestic purchases (USA).

Customer Perspective:

The third page, *Customers*, tracks two primary customer goals; Net Promoter Score and complaints ratio, which is a ratio between the amount of orders where a customer has given a complaint and the total amount of orders.

On the left side, the gauge reflects current NPS, and informs of the amount and share of detractors, passives and promoters. Below, a pistol chart reflects the NPS grouped by human sales and internet sales, while a calendar offers a look into NPS development over time. On the right side, the complaint ratio is shown for each sales person and market.

Figure 32: Customer Scorecard in Power BI

Business Conclusion:

Although performing relatively well on the NPS, the customer satisfaction has been decreasing constantly over time. The customer satisfaction seems to increase when being advised by a sales person, compared to internet retail. However, there are two

sales people that show extraordinarily high complaints rates, namely Pak and Chudukatil. Compared to other countries, the US shows a relative low complaints rate; a good result – since it is IMSport's major market.

Internal Processes Perspective:

The fifth page, *internal processes,* tracks employee happiness, R&D Score and breakage rate. On the top from left, employee happiness is in focus, split in operational staff and management. Average margin and average training expenses are shown just below the KPI-tracker. Two scatter plots are shown for the employees, grouped by management and operational staff.

On the bottom from left, average R&D score for products is measured in a gauge while breakage rate and R&D score per product is plotted in a scatter plot, to the right of the gauge. Finally, to the very right, a bar chart is showing the average breakage rate grouped by product group (*proIsInnovative*) and by year.

Figure 33: Internal Processes Scorecard in Power BI

Business Conclusion:

First, management staff seems to be averagely happier than operative staff. For a sales person, there seems to be a negative correlation between years in company and

happiness. The management should investigate further why long-term employees are less motivated. Also, investments in training show high return since they significantly improve the average margin on sales deals. For products, an investment in research and development seems to decrease the likelihood of breakage during the production.

Learning and Growth Perspective:

The fourth page, *learning and growth*, tracks product innovation rate and performance of sales persons in training and not in training.

On the left side, a sidebar shows the product innovation rate; a result of innovative products released over the total number of products released. Furthermore, important measures for the two product categories and sales persons are shown for comparison

In the middle, filters are available, making it possible to select a product release year, focus on sales persons in training or some specific market. Below a map is shown of the US, the biggest market, indicating with color saturation how much sales each state represents.

On the right, a bar chart shows the relative order quantity of innovative products sold by sales channel and a tree map showing turnover of innovative products by years. If a more detailed view is desired, filter options are useful.

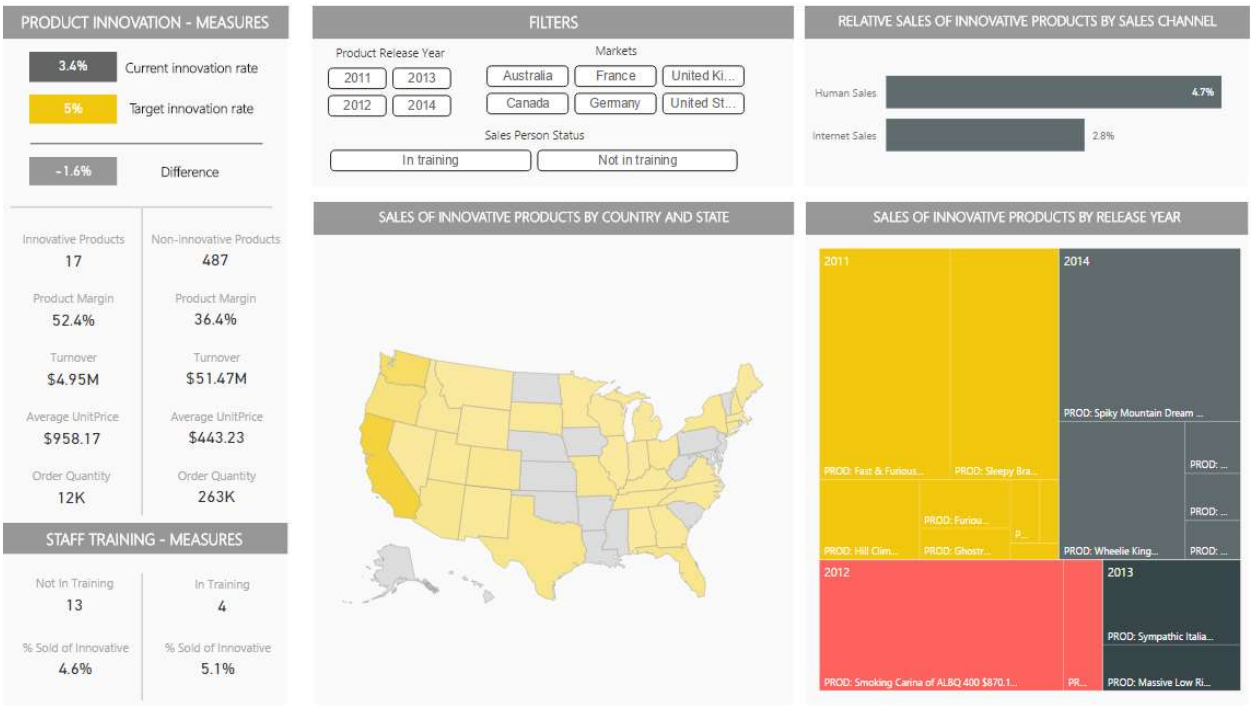


Figure 34: Learning & Growth Scorecard in Power BI

Business Conclusion:

It is revealed that innovative products have a far higher margin and unit price than non-innovative products. Moreover, it is shown that sales persons in training currently have a higher share of sold innovative products, in terms of order quantity. Innovative products are highly demanded at the west coast of the US.

5.5 Bonus Deliverable: R Visualizations in Power BI

To go beyond what was shown in class, we built an additional view in Power BI that is entirely based on the statistical programming language R. Fortunately, there is an R script visual node that supports the integration of R scripts into the dashboard. Four visualization were designed that demonstrate the acquired knowledge of sourcing, manipulating and visualizing data in R:

1. Forecast of monthly profit;
2. Forecast of monthly Net Promoter Score;
3. Analysis of temporal Google trend data;
4. Analysis of geospatial Google trend data.

This will help the management to improve future business and identify trends on the bicycle market. To enable the visualization, one needs an updated version of R as well as the following installed packages: "forecast", "ggplot2", "ggthemes", "zoo", "ggmap", "fiftystater".

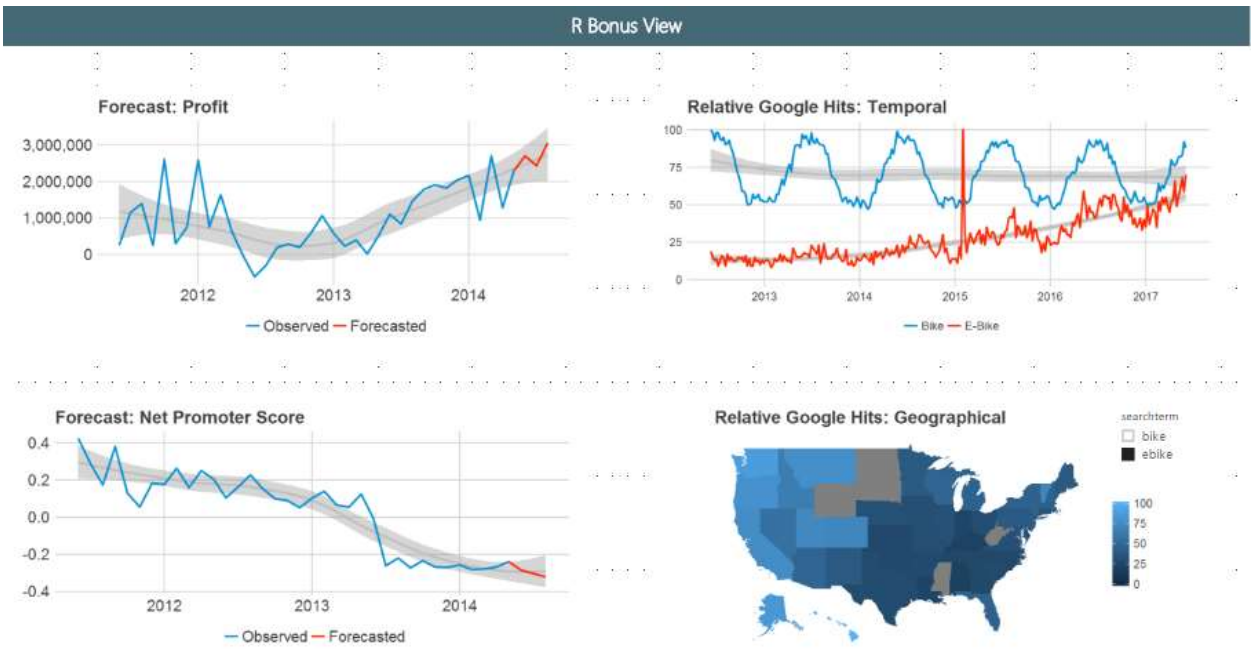


Figure 35: Additional R View in Dashboard

Forecast Visualization:

Initially, tables that contain the foreign key for the time dimension and the measures for profit and net promoter score respectively are created and assigned to the variable *dataset*. The date key is converted to a date format and the forecasted measure is aggregated by month and year. First and last month are excluded to avoid forecasts based on incomplete months.

```
dataset$FK_TimeID <- as.yearmon(as.Date.character(dataset$FK_TimeID,
'%Y%m%d'))

aggr.tbl <- aggregate(SubPriceTotal ~ FK_TimeID, data = dataset, FUN= sum)

aggr.tbl <- aggr.tbl[-c(1,nrow(aggr.tbl)),]
```

Based on the aggregated table, a time series object with monthly frequency was built. An ARIMA model was trained on the series and a forecast for the next three months was created.

```
ts <- ts(aggr.tbl$SubPriceTotal, frequency=12, start = aggr.tbl$FK_TimeID[1])

model <- forecast::auto.arima(ts)

fcast <- forecast::forecast(model, h = 3)
```

A data frame called *fcast.tbl* was created that inherits a set of structural preparation task (not shown) to make us of the data for visualization. For the final visualization, the popular *ggplot2* package in combination with *ggthemes* was used (shortened version):

```
theme_set(theme_fivethirtyeight())

ggplot(data = fcast.tbl, aes(x = date, y = value)) +

  geom_smooth(colour = 'Grey') +

  geom_line(stat = 'identity', size = 1, aes(colour = measure)) +

  scale_y_continuous(labels = scales::comma, limits = c(0,
  max(fcast.tbl$value))) +

  scale_colour_fivethirtyeight() +

  labs(title = 'Revenue Forecast')
```

Business Conclusion:

After yielding in a loss in the middle of 2012, monthly profit has shown a healthy growth with some turbulent ups and downs at the beginning of 2014. The forecast predicts that profit will continue to rise over the next quarter. Also, it shows that the monthly profit will become more stable – a good result. However, looking at the NPS forecast, one can see the opposite pattern. The NPS has been decreasing constantly with a major drop in the middle of 2013. It is forecasted that the NPS will continue to worsen over the course of the next three months. Management should be alarmed about this trend, should investigate what happened in the middle of 2013 and start initiatives to improve customer satisfaction. If this issue is not addressed immediately, future profit outlooks will be likely to diminish.

Trend Analysis:

Google trend data of request for the search terms "bike" and "ebike" were sourced, using the *gtrendsR* package as follows. Unfortunately, Nova IMS internet security standards block the required port, so the data was inserted manually.

```
gconnect(user = "", psw = "")
bike_trend.df <- gtrends(query = "bike", geo = "US", type = "trend")
ebike_trend.df <- gtrends(query = "ebike", geo = "US", type = "trend")
bike_map.df <- gtrends(query = "bike", geo = "US", type = "geo")
ebike_map.df <- gtrends(query = "bike", geo = "US", type = "geo")
```

The data was prepared and finally merged to a variable called *dataset* (not illustrated) The temporal analysis visualization follows the same structure as previously shown code samples. The geographical analysis was visualized, using the *ggmap* and *fiftystater* package:

```
ggplot(dataset, aes(map_id = tolower(state))) +
  geom_map(aes(fill = hits), map = fifty_states) +
  expand_limits(x = fifty_states$long, y = fifty_states$lat) +
  coord_map() +
  scale_x_continuous(breaks = NULL) +
  scale_y_continuous(breaks = NULL) +
  labs(title = 'Trend Analysis: Geographical')
```

Business Conclusion:

The temporal analysis of search terms shows that the popularity of bikes shows a slight downwards trend over time with a strong seasonal pattern that shows peaks in summer and valleys in winter. There variance is constant and there are no outliers. In comparison, e-bikes show a clear increasing popularity with a less significant seasonality. There is a strong outlying value at the beginning of 2015 and the variance seems to increase. This helps the management to understand that electronically powered bicycles are gaining in popularity with no sign of stagnation. E-bikes are frequently searched for in the western part of the US and Hawaii. Launching a new series of e-bikes and promoting it with focus in these states seems to be a promising next step for IMSports.

6 Conclusions & Lessons Learned

Taking advantage of the Microsoft data pipeline is not only useful to manipulate and transform transactional data stored in relational databases but, also, it provides a wide variety of features to perform data analysis and provide insights, always keeping information into context and considering an enterprise-wide business intelligence approach, from both operational and strategic standpoints. The combination of applications used during the execution of the project enabled integration of back-end infrastructure (SQL Server and SSAS) with a front-end dashboard (SSRS and PowerBI) while still allowing users to customize and extract data *ad-hoc* to fulfill their needs (Excel, PowerPivot).

Balanced Scorecard is as powerful tool for defining a managerial strategy and creating relevant measures to evaluate the overall performance of a business unit, especially when combined with the analytical capabilities deployed in this project.

Based on the measures proposed and created with the Balanced Scorecard, we learned that IMSports had a very positive performance over the past years, however, it might be necessary to increase their focus on customer satisfaction.

7 Appendix

7.1 Explanation of measures

- Gross Margin: Turnover subtracted with direct product cost
- Net Promoter Score (NPS): Customer satisfaction measure
- Complains Ratio: Orders with complains/total orders
- Product Innovation Rate: Innovative products released/total products released
- R&D Score: A research score given by extern bureau per product
- Breakage Rate: Number of products that are returned due to breaks