FACULDADE DE ENGENHARIA
Universidade do Porto
Master Data Science and Engineering

HEITOR LIRA
GABRIEL COPOLECCHIA CARVALHAL

Data Warehouse Project

Group 08: Restaurante Olimpia

Porto, Portugal
2022

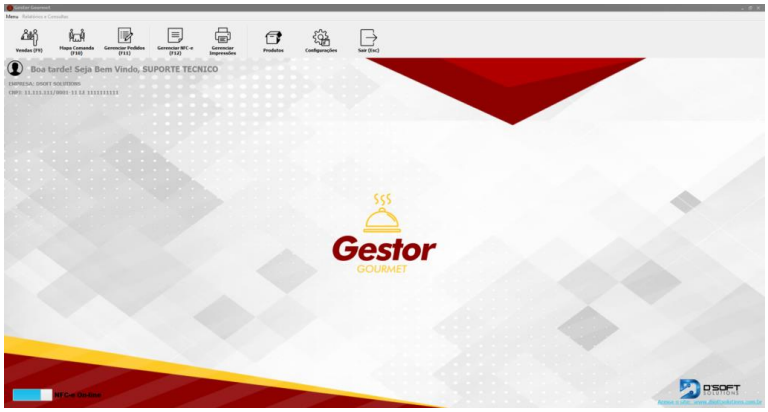
1. Subject description.

The data warehouse project is about Restaurant Olimpia, which is a Brazilian commerce with more than 300 options in your menu. The initial database contains the menu, drinks, dishes recipe, supplies and sales. This structured data will provide the development of queries for data exploration and strategic analysis of the sales and inventory.

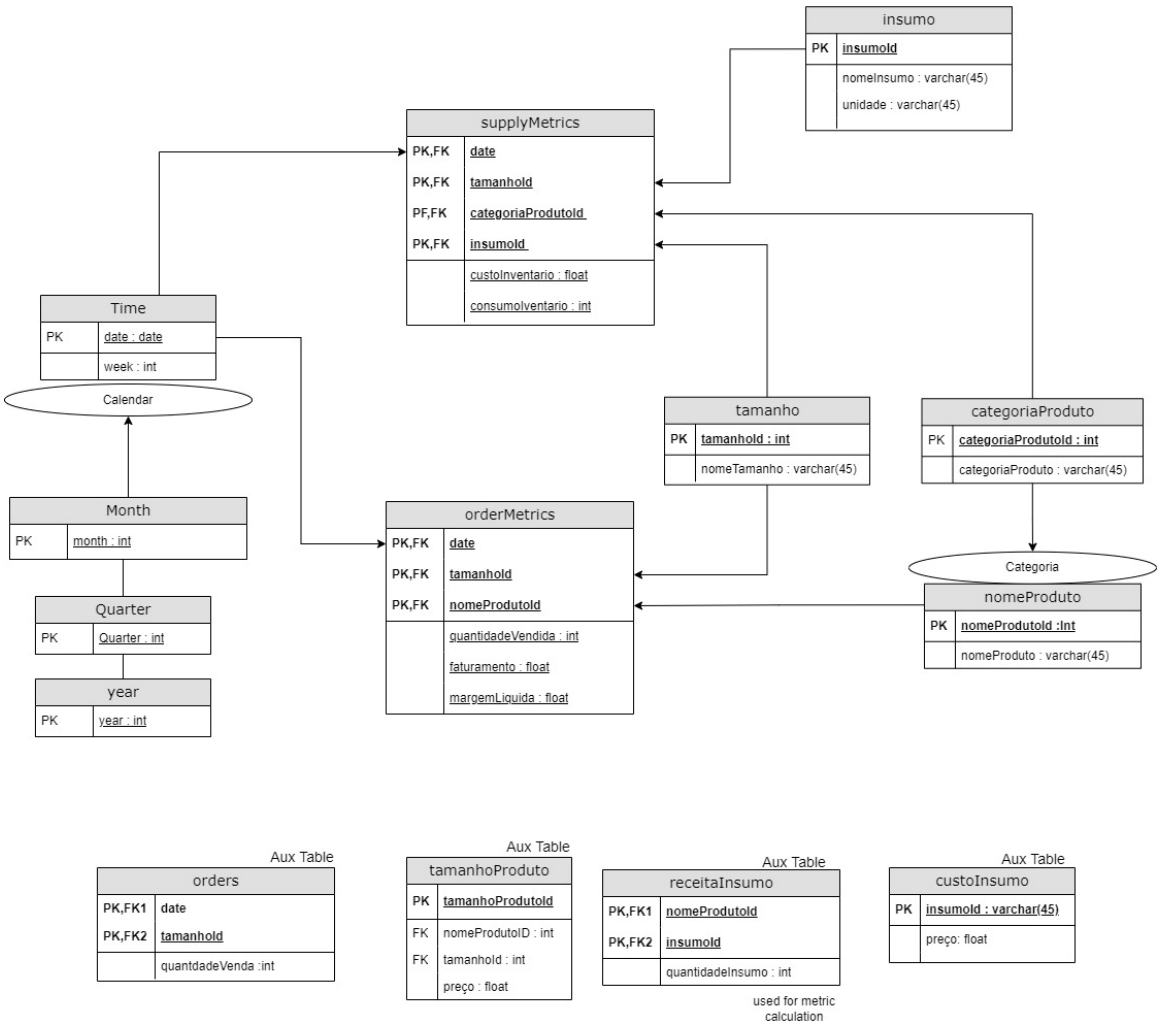
Olimpia Restaurant

Location

ERP



2. Planning: dimensional bus matrix, dimensions, and facts dictionary



Dimension Tables:

- nomeProduto

nomeProduto	fields	Description
	nomeProdutoId	Identifier for product
	nomeProduto	Name of product

- categoriaProduto

categoriaProduto	fields	Description
	categoriaProdutoId	Identifier for product category
	categoriaProduto	Name of product category

- tamanho

tamanho	fields	Description
	tamanhoId	size id for product category
	tamanho	size of product category

- time

Time	fields	Description
	day	Day of the complete date
	week	week of the complete date
	month	Month of the complete date
	quarter	quarter of the complete date
	year	Year of the complete date

- insumos

insumos	fields	Description
	insumoId	Identifier for supplies
	nomeInsumo	supplies name
	unidade	unit of the supplies

Fact Tables:

- orderMetris

orderMetrics	fields	Description
	quantidadeVendida	quantity of products sold
	faturamento	product quantity multiplied by price for each product per order
	margem líquida	(price - cost)/cost

- supplyMetrics

supplyMetrics	fields	Description
	CustoInventário	sum of cost os cost multiplyied by sold quantity
	Consumo inventario	total of supplies used

Auxiliar Table:

- Orders

orders	fields	Description
	tamanhoprodutoId	sales id
	date	first dayof the week
	quantidadeVenda	quantity of sales

- receitaInsumo

receitaInsumo	fields	Description
	quantidadeInsumo	quantity of supplies for each product

- tamanhoProduto

tamanhoProduto	fields	Description
	preço	price of the products

- custoInsumo

custoInsumo	fields	Description
	preço	price of the inputs

3. MultiDim Conceptual model

The subject of the assignment works with two additive measures. Kimball group defines additive as the most flexible and useful facts, because those measures can be summed across any of the dimensions associated with the fact table.

BUSINESS PROCESS		Categoria	Categoria | Produto	Tamanho	Insumo	Calendar
Quantidade vendida	additive		×	×		×
Faturamento	additive		×	×		×
Custo iventário	additive	×		×	×	×
Consumo inventario	additive	×		×	×	×
Margem Líquida	non-additive		×	×		×

Also, the project works with non-additive, such as “Margem Líquida”, which is a ratio between “Faturamento” and “Custo Inventario”. Based on Kimball Group, ratios are non-additive measures, which store the fully additive components of the non-additive measure and sum these components into the final answer set before calculating the final non-additive fact. This final calculation is often done in the BI layer or OLAP cube.

4. Dimensional data model

categoriaProduto (**categoriaProdutoId** , categoriaProduto)

nomeProduto (**nomeProdutoId**,nomeProduto)

tamanho (**tamanhoId**, nomeTamanho)

insumo (**insumoId**, nomeInsumo, unidade)

calendar (**date**, day, month, year, quarter, week, weekDay)

supplyMetrics(**date->calendar**, **insumoId->Insumo**, **tamanhoId->tamanho**, **categoriaId->categoriaProduto,** custoIventario, consumoIventario)

orderMetrics(**date->calendar**,**tamanhoId->tamanho,nomeProdutoId->nomeProduto,** quantidadeVendida, faturamento, margemLiquida)

5. Data sources

Mostly data sources were exported from Olimpia's ERP (Enterprise Resource Planning), as CategoriaProduto, NomeProduto, ReceitaInsumo, Insumos, preço and vendas.

CategoriaProduto, contains product category data, as productID, product name and your category.

nomeProdutoIdOld	nomeProduto	categoria
52	SALADA SIMPLES	SALADAS
153	CAMBUCU CAMARAO ESPECIAL	PEIXES
170	COUVERT A OLIMPIA	COUVERT
172	TORRADA C/ ALHO	COUVERT
173	PATE ATUM	COUVERT
174	MAIONESE ATUM	MAIONESES
175	MAIONESE ATUM	MAIONESES
176	MAIONESE ATUM	MAIONESES
177	MAIONESE FRANGO	MAIONESES
178	MAIONESE FRANGO	MAIONESES
179	MAIONESE FRANGO	MAIONESES
180	MAIONESE CAMARAO	MAIONESES
181	MAIONESE CAMARAO	MAIONESES
182	MAIONESE CAMARAO	MAIONESES
183	MAIONESE LEGUMES	MAIONESES
184	MAIONESE LEGUMES	MAIONESES
185	MAIONESE LEGUMES	MAIONESES
186	MAIONESE CAMARAO ESPECIAL	MAIONESES
187	MAIONESE CAMARAO ESPECIAL	MAIONESES

NomeProduto, contains product data, as productID, product name, size, and different information about category. In this case, categoriaProduto distinguish products between receitas(recipes) and bebidas (beverages), which is irrelevant to the model and in the transformation process will be ignored.

nomeProdutoIdOld	categoriaProduto	nomeProduto	tamanh
52	receitas	SALADA SIMPLES	MEIO
153	receitas	CAMBUCU CAMARAO ESPECIAL	MEIO
170	receitas	COUVERT A OLIMPIA	ÚNICO
172	receitas	TORRADA C/ ALHO	ÚNICO
173	receitas	PATE ATUM	ÚNICO
174	receitas	MAIONESE ATUM	INTEIRO
175	receitas	MAIONESE ATUM	MEIO
176	receitas	MAIONESE ATUM	INDIVIDUAL
177	receitas	MAIONESE FRANGO	INTEIRO
178	receitas	MAIONESE FRANGO	MEIO
179	receitas	MAIONESE FRANGO	INDIVIDUAL
180	receitas	MAIONESE CAMARAO	INTEIRO
181	receitas	MAIONESE CAMARAO	MEIO
182	receitas	MAIONESE CAMARAO	INDIVIDUAL
183	receitas	MAIONESE LEGUMES	INTEIRO
184	receitas	MAIONESE LEGUMES	MEIO
185	receitas	MAIONESE LEGUMES	INDIVIDUAL
186	receitas	MAIONESE CAMARAO ESPECIAL	INTEIRO
187	receitas	MAIONESE CAMARAO ESPECIAL	MEIO

ReceitaInsumo, contains the detailed data about recipes and beverages, how much each product receives from each ingredient.

Codigoprato	Codigoinsumo	Quantidadeinsun
52	L100000	150
52	L100001	150
52	L100002	100
153	A100000	150
153	A100001	500
153	A100005	50
153	A100041	10
153	L100001	600
153	L100002	200
153	L100009	100
153	L100014	100
153	M100001	400
153	M100011	500
170	A100002	100
170	A100003	100
170	A100004	3
170	A100005	20
170	A100006	10
170	L100003	1000

Insumos, contains the detailed information about all the ingredients bought by the restaurant. The table hac input ID, input name, unit of the ingredient and price per unit of the input.

Codigoinsumo	Insumo	Unidade	Custo
A100000	ARROZ BRANCO	Gramas	10,29
A100001	BATATA	Gramas	25
A100002	OVO CODORNA	Gramas	57,58
A100003	AZEITONA PRETA	Gramas	82,59
A100004	PÃO FRANCÊS	Gramas	0,0636
A100005	ALHO	Gramas	0,02
A100006	MAIONESE	Gramas	0,02702
A100007	MUSSARELA	Gramas	0,03375
A100008	OVO	Unidade	1,105
A100009	PROVOLONE	Gramas	0,03875
A100010	LEITE DE COCO	Mililitros	0,01645
A100011	AZEITE DENDÊ	Mililitros	0,021877778
A100012	FARINHA DE MANDIOCA	Gramas	0,00566
A100013	FARINHA DE ROSCA	Gramas	0,0108
A100014	FARINHA DE TRIGO	Gramas	0,00457
A100015	COLORAL	Gramas	0,1366
A100016	CALDO DE GALINHA	Tablete	0,63
A100017	FEIJÃO PRETO	Gramas	0,00866
A100018	QUEIJO PRATO	Gramas	0,03257

Preço, contains the price of all products offered by olimpia, as product Id, product name an the price.

GTIN	NOME DO PRODUTO	**PREÇO**	
210	CASQUINHA DE SIRI	R$	24,00
43	ORIGINAL 600ML	R$	15,00
267	CAMARAO A GREGA (MEIO)	R$	132,00
34	SUCO DE FRUTA (JARRA)	R$	20,00
417	FILE DE PESCADA A MODA DA CASA (MEIO)	R$	106,00
33	SUCO DE FRUTA (COPO)	R$	11,00
282	CAMARAO AO CATUPIRY ESPECIAL (MEIO)	R$	162,00
45	HEINEKEN 600ML	R$	15,00
266	CAMARAO A GREGA (INTEIRO)	R$	223,00

Vendas, weekly sales aggregated by product. The table has, date, product ID, product name, quantity sold and the invoicing per week and product. Desc and Lucro columns are ignored by the model in the transformation process.

DataBase	GTIN	NOME DO PRODUTO	QTDADE	VLR TOTAL	DESC	LUCRO
09/10/2021	1	AGUA S/ GAS	107	856	0	856
09/10/2021	2	AGUA C/ GAS	85	680	4,4	675,6
09/10/2021	3	COCA LATA	50	450	0	450
09/10/2021	4	COCA ZERO LATA	46	414	0	414
09/10/2021	5	FANTA LAR LATA	4	36	0	36
09/10/2021	7	GUARANA LATA	40	360	0	360
09/10/2021	8	GUARANA ZERO LATA	23	207	0	207
09/10/2021	9	TONICA LATA	26	234	0	234
09/10/2021	10	TONICA ZERO LATA	7	63	0	63
09/10/2021	12	SODA LATA	7	63	0	63
09/10/2021	13	SODA ZERO LATA	4	36	0	36
09/10/2021	14	H2OH LIMAO	24	264	0,01	263,99
09/10/2021	15	H2OH LIMONETO	9	99	0	99
09/10/2021	16	COCA KS	161	1.368,50	1,22	1.367,28
09/10/2021	18	SCHWEPPES CITRUS LATA	16	160	0	160
09/10/2021	19	ICE TEA	2	20	0	20
09/10/2021	20	COCA 600ML	46	506	0,02	505,98
09/10/2021	22	GUARANA 600ML	15	165	0	165
09/10/2021	24	SODA 600ML	4	44	0	44

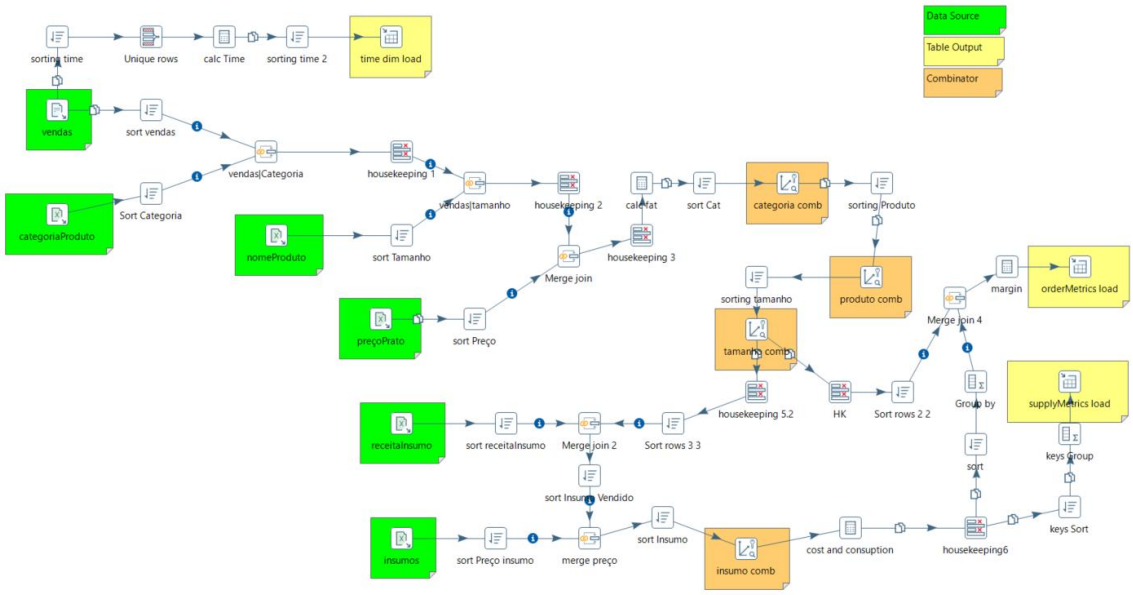
6. Transformations

The ETL process was carried out from the individual import of the aforementioned tables and the application of a series of joins to assemble the necessary bases to perform the export of the necessary fact and dimensions tables.

All dimension tables, except for time Dimension, were created using combinator operations.This decision was made due to the fact that the column (date) of the original table was kept as a primary key for this dimension, instead of adding a new surrogateKey, a step performed inside the combinator operation by default.

the sources "vendas", "categoriaProduto", "nomeProduto" and “precoPrato" were initially used to calculate the quantity sold of each product, making it possible to calculate the revenue. Then, the datastream is splitted in two. The first one is used to, using the quantity sold of each product, identify which and how many ingredients were used for each of the records (information obtained by merging with the “receitaInsumo” table). Then, the “insumos” table, which contains the market price of each of the listed ingredients, is merged to allow the calculation of the total cost. This table is finally sorted and exported to the database as the supplyMetrics fact.

The second data stream is later used to be merged with the product of the latest (grouped by the product) to allow the calculation of gross margins for each record. This table is then exported as the OrderMetrics fact table

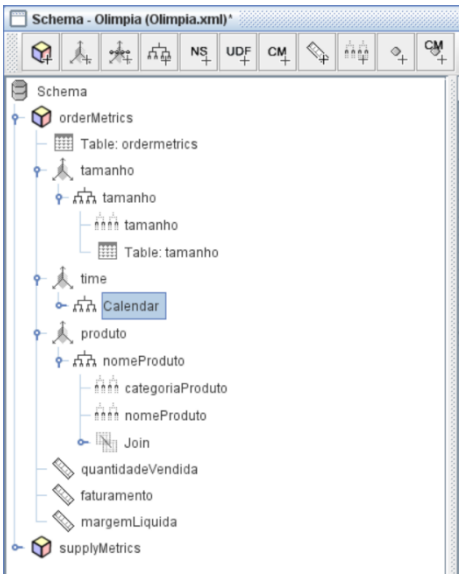


7. Multi-dimensional modeling

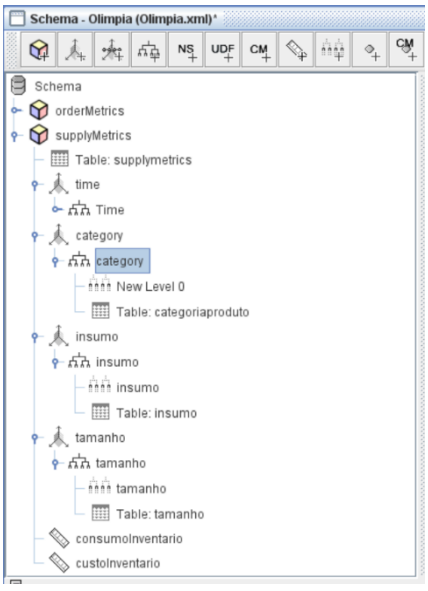
Although the data modeling were firstly designed to be an Constalation Schema, with 2 fact tables, during the creation of the cubes, the group decided to create, instead, to build 2 different schemas: One cube as a star schema,using SupplyMetric as Fact and insumo,categoriaProduto, tamanho and time as dimensions, and ;one cube with a snowflake schema,with orderMetrics as fact table and insumo, nomeProduto, categoriaProduto, tamanho and time as dimensions. CategoriaProduto , in this case, is being connect to nomeProduto instead of connected to the orderMetrics directly.

The Mondrian Cubes created using the schema workbench are shown bellow and the XML generated is attached with this report.

orderMetrics Cube



supplyMetrics Cube

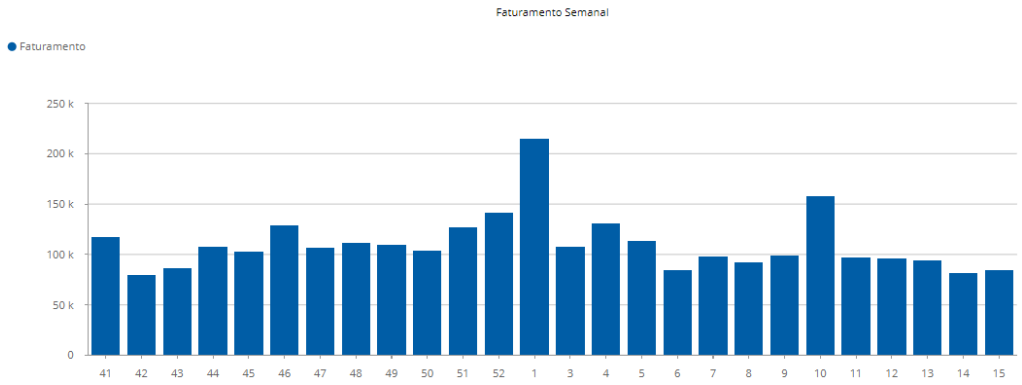


8. Data analysis (dashboards and MDX queries)

8.1. PENTAHO OLAP

The following graphs were generated using Pentaho OLAP server and the underlining MDX query specified.

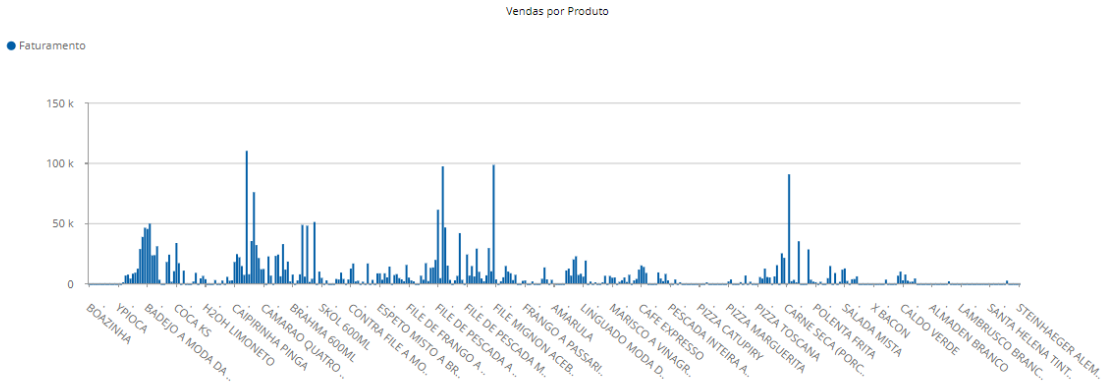
8.1.1. Weekly Revenue



MDX Query:

```
SELECT
    NON EMPTY({Descendants([Time].[All Times] ,[Time].[Week])}) on ROWS,
    NON EMPTY({[Measures].[Faturamento]}) on Columns
FROM [orderMetrics]
```

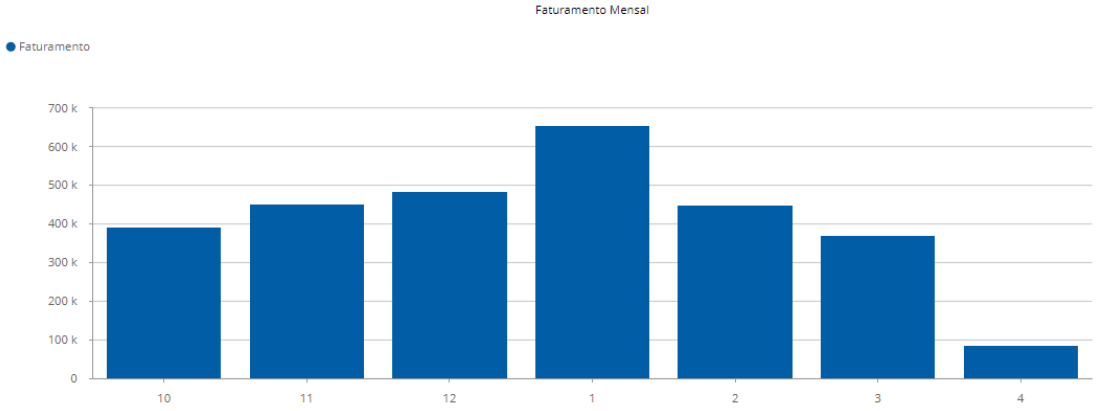
8.1.2. Revenue by Product



MDX Query:

```
SELECT
    NON EMPTY({Descendants([Nomeproduto].[All Nomeprodutos] ,[Nomeproduto].[NomeProduto])}) on ROWS,
    NON EMPTY({[Measures].[Faturamento]}) on Columns
FROM [orderMetrics]
```

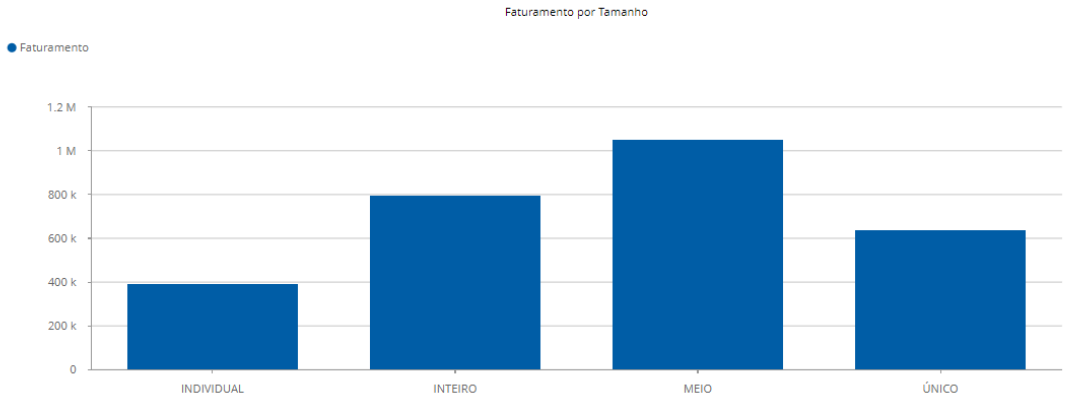
8.1.3. Montly Revenue



MDX Query:

```
SELECT
    NON EMPTY({Descendants([Nomeproduto].[All Nomeprodutos] ,[Nomeproduto].[NomeProduto])}) on ROWS,
    NON EMPTY({[Measures].[Faturamento]}) on Columns
FROM [orderMetrics]
```

8.1.4. Revenue by dish size

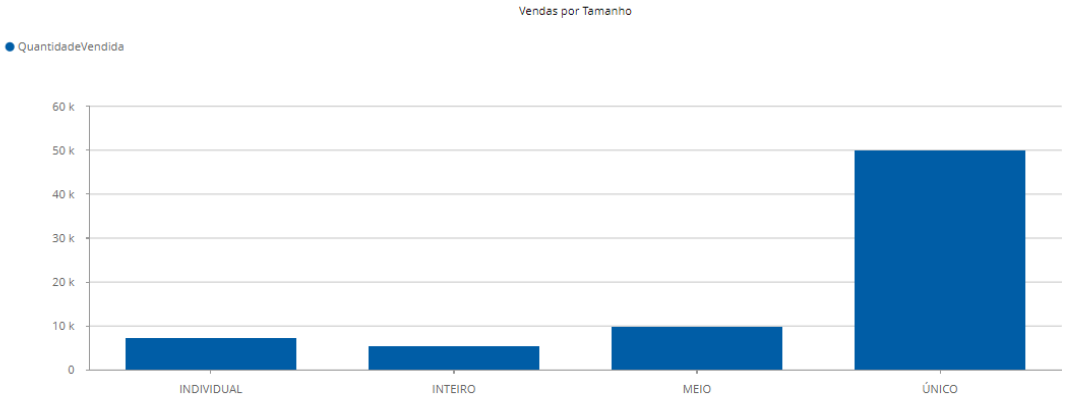


MDX Query:

```
SELECT
   NON EMPTY({Descendants([Tamanho.NomeTamanho].[All
Tamanho.NomeTamanhos],[Tamanho.NomeTamanho].[NomeTamanho])}) on ROWS,
    NON EMPTY({[Measures].[Faturamento]}) on Columns

FROM [orderMetrics]
```

8.1.5. Sales Number by dish size

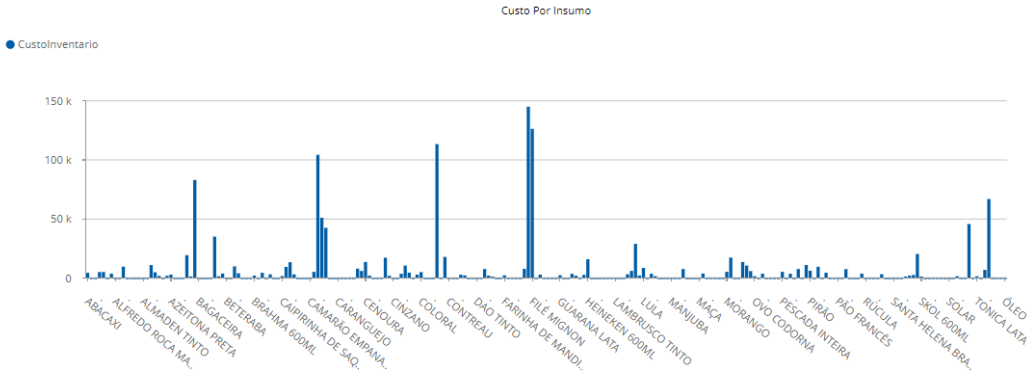


MDX Query:

```
SELECT
    NON EMPTY({Descendants([Tamanho.NomeTamanho].[All
    Tamanho.NomeTamanhos],[Tamanho.NomeTamanho].[NomeTamanho])}) on ROWS,
    NON EMPTY({[Measures].[QuantidadeVendida]}) on Columns

FROM [orderMetrics]
```

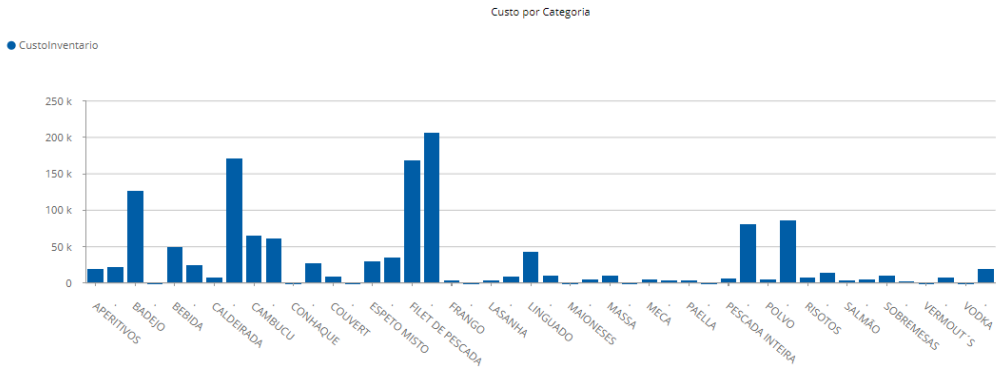
8.1.6. Input Costs



MDX Query:

```
SELECT
      NON EMPTY({Descendants([Insumo.NomeInsumo].[All Insumo.NomeInsumos]
,[Insumo.NomeInsumo].[NomeInsumo])}) on ROWS,
      NON EMPTY({[Measures].[CustoInventario]}) on Columns
 FROM [supplyMetrics]
```

8.1.7. Cost by Category

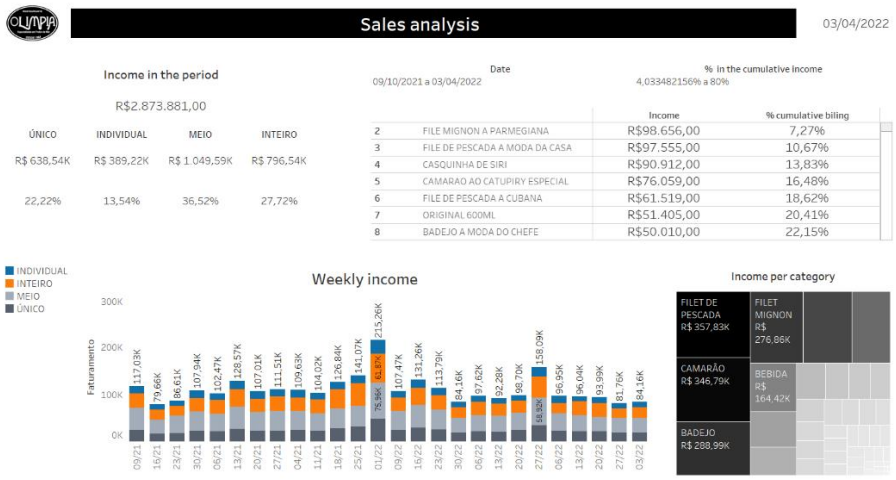


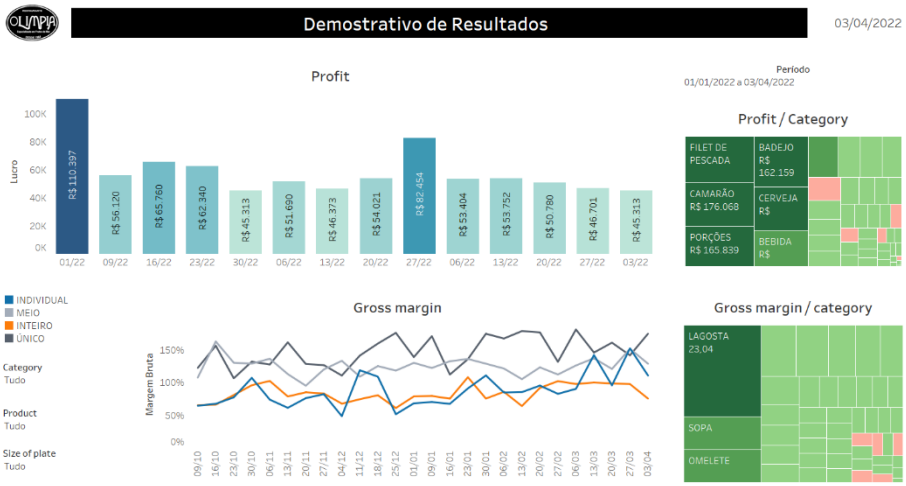
MDX Query:

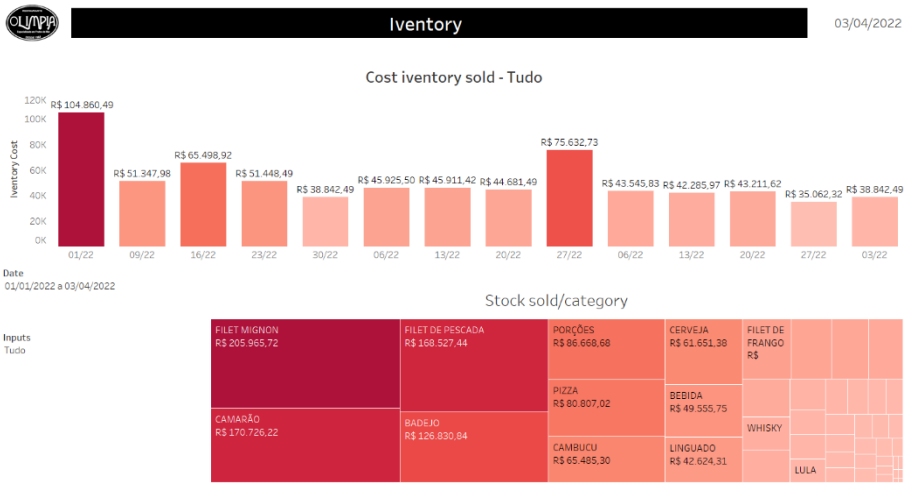
```
select NON EMPTY({Descendants([Categoriaproduto.NomeCategoria].[All Categoriaproduto.NomeCategorias]
,[Categoriaproduto.NomeCategoria].[NomeCategoria])}) on ROWS,
 NON EMPTY({[Measures].[CustoInventario]}) on Columns
 from [supplyMetrics]
```

8.2. Tableau DashBoard

Due to the limitations found by the group to perform all the analysis desired in the Pentaho OLAP Server,it was decided to migrate the schemas to Tableau and create an additional DashBoard using the tool. The results are presented bellow.







9. Conclusion

Using the views generated by the Pentaho OLAP server and Tableau, it was possible to draw a series of conclusions about the performance of the Olimpia restaurant, some of which are not obvious and beneficial to the owner.

As expected, the last and first weeks of the year (summer in the Southern Hemisphere) are the peak seasons, with another peak at the end of February, due to Carnival, which is very popular in Brazil.

The analysis of the behavior of the margin by plate size also presented results in line with expectations, with sales of half portions showing better margins than whole ones. However, lower volatility in the margin of individual size products was expected.

The presence of some product categories that presented negative margin ,such as liquors, Whiskys, appetizers and pizzas, lights up a red light for the review of prices, or

even the necessity of discontinue some products offered, in order to improve the overall margin of the establishment.

10. Attachments

10.1. Database Create Script

```
## Create Tables

# CategoriaProduto
CREATE TABLE `categoriaproduto` (
  `categoriaProdutoId` int(11) NOT NULL,
  `nomeCategoria` varchar(45) NOT NULL,
  PRIMARY KEY (`categoriaProdutoId`)
);

#Insumo
CREATE TABLE `insumo` (
  `insumoId` int(11) NOT NULL,
  `nomeInsumo` varchar(45) DEFAULT NULL,
  `unidade` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`insumoId`)
);

#nome produto
CREATE TABLE `nomeproduto` (
  `nomeProdutoId` int(11) NOT NULL,
  `categoriaProdutoId` int(11) NOT NULL,
  `nomeProduto` varchar(45) NOT NULL,
  PRIMARY KEY (`nomeProdutoId`),
  FOREIGN KEY (`categoriaProdutoId`) REFERENCES categoriaproduto(
`categoriaProdutoId`)
);

#tamanho
CREATE TABLE `tamanho` (
  `tamanhoId` int(11) NOT NULL,
  `nomeTamanho` varchar(45) NOT NULL,
  PRIMARY KEY (`tamanhoId`)
);

#time
CREATE TABLE `time` (
  `date` date NOT NULL,
  `month` int(11) NOT NULL,
  `week` int(11) NOT NULL,
  `quarter` int(11) NOT NULL,
  `year` int(11) NOT NULL,
  PRIMARY KEY (`date`)
);


#order metrics
CREATE TABLE `ordermetrics` (
  `date` date NOT NULL,
  `tamanhoId` int(11) NOT NULL,
  `nomeProdutoId` int(11) NOT NULL,
```

```
  `quantidadeVendida` int(11) NOT NULL,
  `faturamento` decimal(10,0) NOT NULL,
  `margemLiquida` float NOT NULL,
  PRIMARY KEY (`date`,`tamanhoId`,`nomeProdutoId`),
  FOREIGN KEY (`nomeProdutoId`) REFERENCES nomeproduto(
`nomeProdutoId`),
  FOREIGN KEY (`date`) REFERENCES time( `date`),
  FOREIGN KEY (`tamanhoId`) REFERENCES tamanho( `tamanhoId`)
);

#supply metrics
CREATE TABLE `supplymetrics` (
  `date` date NOT NULL,
  `tamanhoId` int(11) NOT NULL,
  `categoriaProdutoId` int(11) NOT NULL,
  `insumoId` int(11) NOT NULL,
  `custoInventario` float NOT NULL,
  `consumoInventario` float NOT NULL,
  PRIMARY KEY (`date`,`tamanhoId`,`categoriaProdutoId`,`insumoId`),
  FOREIGN KEY (`categoriaProdutoId`) REFERENCES categoriaproduto(
`categoriaProdutoId`),
  FOREIGN KEY (`date`) REFERENCES time( `date`),
  FOREIGN KEY (`tamanhoId`) REFERENCES tamanho( `tamanhoId`),
  FOREIGN KEY (`insumoId`) REFERENCES insumo(`insumoId`)
);
```

10.2. DATAWAREHOUSE DIAGRAM