NAME : ARUL KUMAR ARK

ROLLNO :225229103

Department of Data Science Bishop Heber College Tiruchirappalli

NoSQL Database Management Lab

Lab4. Retail Sales Analytics Part I

Objectives

In this lab, you will be performing analytics on retail sales of a particular enterprise. The sales of several of their stores are represented using three excel flies.

The Problem

One challenge of modelling retail data is the need to make decisions based on limited history. Holidays and select major events come once a year, and so does the chance to see how strategic decisions impacted the bottom line. In addition, markdowns are known to affect sales – the challenge is to predict which departments will be affected and to what extent.

Data Description

You are provided with historical sales data for 45 stores located in different regions each store contains a number of departments. The company also runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks.

Attributes of Stores, Features and Sales Tables

1. Stores

Anonymized information about the 45 stores, indicating the type and size of store

2. Features

Contains additional data related to the store, department, and regional activity for the given dates.

- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel Price - cost of fuel in the region
- MarkDown1-5 - anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

3. Sales

Historical sales data, which covers to 2010-02-05 to 2012-11-01. Within this tab you will find the Following fields:

- Store - the store number
- Dept - the department number
- Date - the week
- Weekly_Sales - sales for the given department in the given store
- IsHoliday - whether the week is a special holiday week

Some Sample Tasks

1. Predict the department-wide sales for each store for the following year
2. Model the effects of markdowns on holiday weeks
3. Provide recommended actions based on the insights drawn, with prioritization placed on largest business impact

Tasks To be Completed:

Question1: Describe the data, making note of the interesting attributes and relationships in the data. Include 3-5 rows of sample data from each file. Very important: explain what insights you hope to gain from analyzing the data.

SQL> desc stores;

Name	Null?	Type
STORE_ID		NUMBER(38)
TYPE		VARCHAR2(26)
SIZE_		NUMBER(38)

SQL> desc features;

Name	Null?	Type
STORE_ID		NUMBER(38)
DATE_		DATE
TEMPERATURE		NUMBER(38,2)
FUEL_PRICE		NUMBER(38,3)
MARKDOWN1		NUMBER(38,2)
MARKDOWN2		NUMBER(38,2)
MARKDOWN3		NUMBER(38,2)

MARKDOWN4	NUMBER(38,2)
MARKDOWN5	NUMBER(38,2)
CPI	NUMBER(38,7)
UNEMPLOYMENT	NUMBER(38,3)
ISHOLIDAY	VARCHAR2(26)

SQL> desc sales;

Name	Null?	Type
SALES_ID		NUMBER(38)
STORE_ID		NUMBER(38)
DEPT		NUMBER(38)
DATE_		DATE
WEEKLY_SALES		NUMBER(38,2)
ISHOLIDAY		VARCHAR2(26)

Question2: Create tables from Excel files. Relate the tables via primary key and foreign key wherever required.

alter table stores add primary key(store_id);

alter table features add foreign key(store_id) references stores(store_id);

alter table sales add foreign key(store_id) references stores(store_id);

Question3: Write atleast 2 queries to explore each table by using WHERE and ORDERBY

select * from sales where sales_id<=5 order by sales_id;

SALES_ID	STORE_ID	DEPT	DATE_	WEEKLY_SALES	ISHOLIDAY
1	1	1	05-FEB-10	24924.5	FALSE
2	1	1	12-FEB-10	46039.49	TRUE
3	1	1	19-FEB-10	41595.55	FALSE
4	1	1	26-FEB-10	19403.54	FALSE
5	1	1	05-MAR-10	21827.9	FALSE

SQL> select * from sales where weekly_sales=20 and store_id=45 order by weekly_sales;

SALES_ID	STORE_ID	DEPT	DATE_	WEEKLY_SALES	ISHOLIDAY
417807	45	47	25-MAR-11	20	FALSE
417808	45	47	08-APR-11	20	FALSE
417809	45	47	13-MAY-11	20	FALSE
417866	45	49	22-JUN-12	20	FALSE
418067	45	54	16-JUL-10	20	FALSE
418107	45	54	20-MAY-11	20	FALSE

SQL> select * from stores where store_id<5 order by store_id;

STORE_ID	TYPE	SIZE_
1	A	151315
2	A	202307
3	B	37392
4	A	205863

SQL> select * from stores where type='A' and store_id<=5 order by type;

STORE_ID	TYPE	SIZE_
1	A	151315
2	A	202307
4	A	205863

SQL> select * from features where store_id=1 and markdown3=1 order by store_id;

STORE_ID	DATE_	TEMPERATURE	FUEL_PRICE	MARKDOWN1	MARKDOWN2	MARKDOWN3	MARKDOWN4	MARKDOWN5	CPI	UNEMPLOYMENT	ISHOLIDAY
1	26-JUL-13	81.54	3.62	332.17	673.19	1	38.63	1245.2			FALSE

SQL> select * from features where isholiday='TRUE' and store_id=45 order by isholiday;

STORE_ID DATE_ TEMPERATURE FUEL_PRICE MARKDOWN1 MARKDOWN2 MARKDOWN3
MARKDOWN4 MARKDOWN5 CPI UNEMPLOYMENT ISHOLIDAY
45 12-FEB-10 27.73 2.773
181.982317 8.992 TRUE
45 10-SEP-10 70.87 2.699
182.598178 8.743 TRUE
45 26-NOV-10 46.15 3.039
182.783277 8.724 TRUE
45 31-DEC-10 29.67 3.179
182.571448 8.724 TRUE
45 11-FEB-11 30.3 3.239
183.701613 8.549 TRUE
45 09-SEP-11 71.48 3.738
186.673738 8.625 TRUE
45 25-NOV-11 48.71 3.492 140.87 384.82 26961.99
28.59 1110.12 188.3504 8.523 TRUE
45 30-DEC-11 37.79 3.389 3604.35 43941.56 287.45
357.43 750.01 189.062016 8.523 TRUE
45 10-FEB-12 37 3.64 15760.3 6113.95 5.43
5656.35 4685.58 189.707605 8.424 TRUE
45 07-SEP-12 75.7 3.911 11024.45 12.8 52.63
1854.77 2055.7 191.577676 8.684 TRUE
45 23-NOV-12 43.08 3.748 1005.79 72542.01
484.7 620.12 192.283032 8.667 TRUE
45 28-DEC-12 35.96 3.563 12751.77 24656.74 230.07
31.46 1654.14 192.559264 8.667 TRUE
45 08-FEB-13 28.99 3.753 53311.88 531.33 78.26
24823.94 3233.44 192.897089 8.625 TRUE