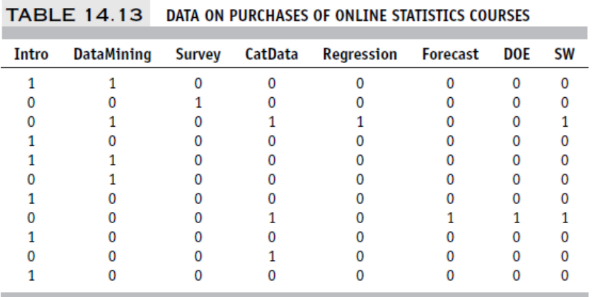
**BI Analyzation 2**



**14.2 Identifying Course Combinations:** The Institute for Statistics Education at Statistics.com

offers online courses in statistics and analytics, and is seeking information that will help in

packaging and sequencing courses. Consider the data in the file Coursetopics.csv, the first

few rows of which are shown in Table 14.13. These data are for purchases of online statistics

courses at Statistics.com. Each row represents the courses attended by a single customer.

The firm wishes to assess alternative sequencings and bundling of courses.

**Use association rules to analyze the data and interpret several of the resulting rules (“Run R”).**

combos = read.csv("Coursetopics.csv")

str(combos)

summary(combos)

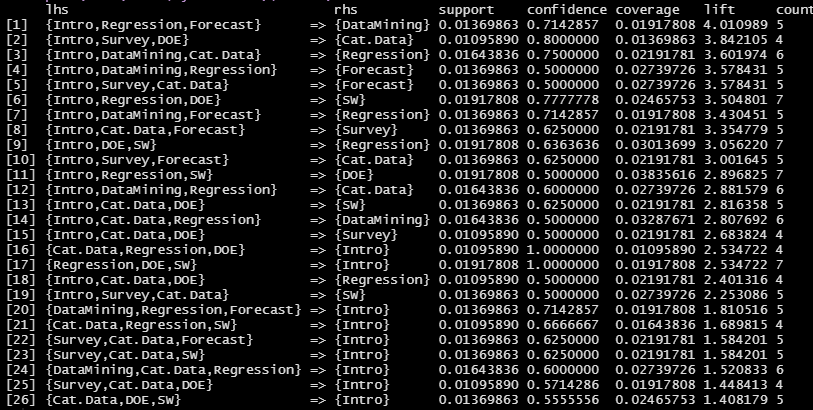
combos.mat<-as.matrix(combos)

combos.trans <- as(combos.mat,"transactions")

rules <- apriori(combos.trans,parameter = list(supp = 0.01, conf = 0.5, minlen = 4))

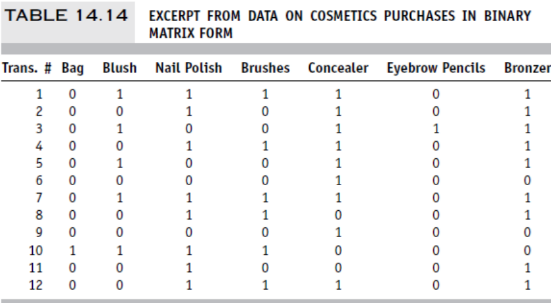
inspect(sort(rules,by = "lift"), n = 6)

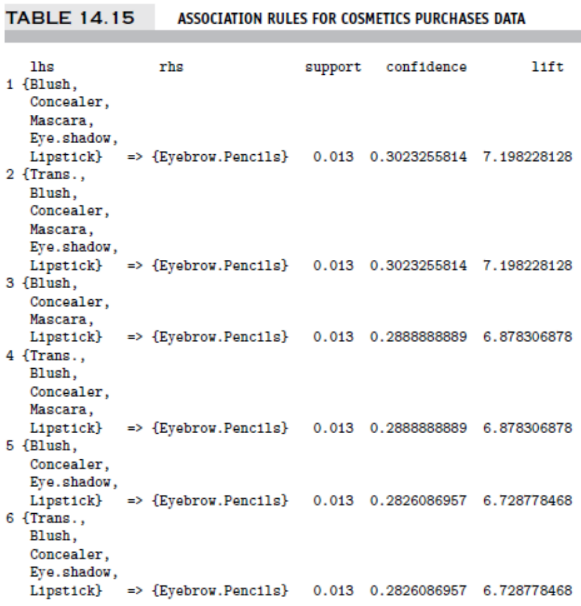
**My R Results: 26 Rules**



**My Interpretations**:

1. Rule [1]: We are 71.4% confident that students will likely choose to take the DataMining course if they are also enrolled in Intro, Regression, and Forecast. Since the lift of 4.01 > 1, we can safely assume these courses are combined frequently.
2. Rule [2]: We are 80.0% confident that students will likely choose to take the Cat.Data course if they are also enrolled in Intro, Survey, and DOE. Since the lift of 3.84 > 1, we can safely assume these courses are combined frequently.
3. Rule [3]: We are 75.0% confident that students will likely choose to take the Regression course if they are also enrolled in Intro, DataMining, and Cat.Data. Since the lift of 3.60 > 1, we can safely assume these courses are combined frequently.





**14.4 Cosmetics Purchases:** The data shown in Table 14.14 and the output in Table 14.15 are

based on a subset of a dataset on cosmetic purchases (Cosmetics.csv) at a large chain

drugstore. The store wants to analyze associations among purchases of these items for

purposes of point-of-sale display, guidance to sales personnel in promoting cross-sales, and

guidance for piloting an eventual time-of-purchase electronic recommender system to boost

cross-sales.

**Consider first only the partial data shown in Table 14.14, given in binary matrix form:**

**(a) Select several values in the matrix and explain their meaning.**

Given the matrix, we notice the following for baskets 1 - 3:

* ***Basket 1***: includes Blush, Nail Polish, Brushes, Concealer, and Bronzer
* ***Basket 2***: includes Nail Polish, Concealer, and Bronzer
* ***Basket 3***: includes Blush, Concealer, Eyebrow Pencils, and Bronzer

**Next, consider the partial output in Table 14.15.**

**(b) Consider the results of the association rules analysis shown in Table 14.15.**

**i. For the first row, explain the “confidence” output and how it is calculated.**

* **First Row Interpretation**:
  + *When customers purchase blush, concealer, mascara, eye shadow, and lipstick, they also purchase eyebrow pencils 30.23% of the time.*
* ***Confidence* Formula**:

*# of transactions containing items in antecedent and consequent*

***Confidence*** = -------------------------------------------------------------------------------------

*# of transactions containing items in antecedent*

**ii. For the first row, explain the “support” output and how it is calculated.**

* **First Row Interpretation**:
  + *In 1.3% of all transactions, customers who purchase blush, concealer, mascara, eye shadow, and lipstick* ***also*** *purchase eyebrow pencils.*
* ***Support*** **Formula**:

*# of transactions containing all items in the antecedent and consequent*

***Support*** = ------------------------------------------------------------------------------------------------

*# of transactions in the database*

**iii. For the first row, explain the “lift” and how it is calculated.**

* **First Row Interpretation**:
  + *Given rule 1, when customers purchase blush, concealer, mascara, eye shadow, and lipstick, the likelihood they also purchase eyebrow pencils increases by 7.198 times.*
* ***Lift* Formula**:

*Confidence of the Rule*

***Lift*** *= -------------------------------------*

*Support of the Consequent*

**iv. For the first row, explain the rule that is represented there in words.**

* When customers purchase blush, concealer, mascara, eye shadow, and lipstick, the same customers will also purchase eyebrow pencils.

**(c) Now, use the complete dataset on the cosmetics purchases (in the file Cosmetics.csv).**

**“Run” association rule using R (e.g., use the default parameters).**

##Cosmetics

cosmoCombos = read.csv("Cosmetics.csv")

str(cosmoCombos)

summary(cosmoCombos)

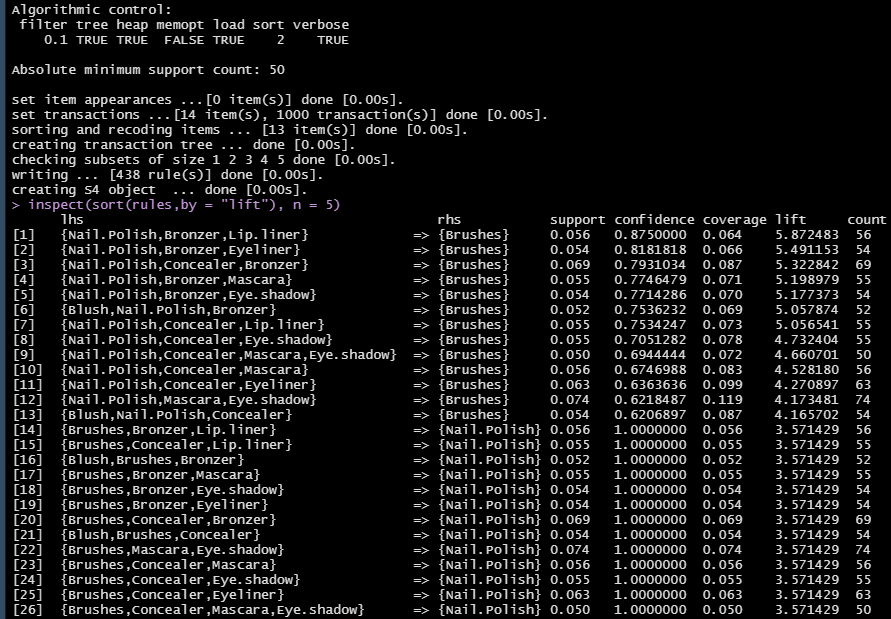
cosmoCombos.mat<-as.matrix(cosmoCombos[,-1])

cosmoCombos.trans <- as(cosmoCombos.mat,"transactions")

rules <- apriori(cosmoCombos.trans,parameter = list(supp = 0.05, conf = 0.5, minlen = 4))

inspect(sort(rules,by = "lift"), n = 5)

**My R Code Results: 438 Rules *(\*Screenshot only displays 26 Rules!)***



**i. Interpret the first three rules in the output in words.**

1. Rule [1]: We are 87.5% confident that customers will likely choose to purchase the Brushes if they are also purchasing Nail Polish, Bronzer, and Lip Liner. Since the lift of 5.87 > 1, we can safely assume these items are purchased together frequently.
2. Rule [2]: We are 81.8% confident that customers will likely choose to purchase the Brushes if they are also purchasing Nail Polish, Bronzer, and Eye Liner. Since the lift of 5.49 > 1, we can safely assume these items are purchased together frequently.
3. Rule [3]: We are 79.3% confident that customers will likely choose to purchase the Brushes if they are also purchasing Nail Polish, Concealer, and Bronzer. Since the lift of 5.32 > 1, we can safely assume these items are purchased together frequently.

**ii. Reviewing the first couple of dozen rules, comment on their redundancy and how you**

**would assess their utility.**

* After sorting the data in descending order by lift and reviewing the first 24 rules, I notice a number of patterns in the association rules:
  + Rows 1-13 are very similar to rows 14-26.
  + Brushes, Nail Polish, and Concealer appear in 7 of the first 13 rows, so they appear to be items which are grouped very frequently.
  + Items in Rows 1-7 have a lift > 5; therefore, customers who purchase the first 3 items on the left-hand side are over 500% more likely to grab Brushes as well.
* This data would be useful in figuring out which items to either cross-merchandise among the different departments of the drugstore.
* This data would also be useful in figuring out what items to include in a specific endcap display.
* This data would also help with deciding what items to “bundle” together.