Dillards: Retail Sales Data Analysis

1. Get to know your data
    a. Determine how many distinct skus there are in pairs of the skuinfo, skstinfo, and trnsact tables. Which skus are common to pairs of tables, or unique to specific tables?
    b. Determine how many instances there are of each sku associated with each store in the skstinfo table and the trnsact table?
    c. Determine how many distinct stores there are in the strinfo, store_msa, skstinfo, and trnsact tables.
    d. Which stores are common to all four tables, or unique to specific tables?
    e. It turns out there are many skus in the trnsact table that are not in the skstinfo table. As a consequence, we will not be able to complete many desirable analyses of Dillard's profit, as opposed to revenue, because we do not have the cost information for all the skus in the transact table. Examine some of the rows in the trnsact table that are not in the skstinfo table; can you find any common features that could explain why the cost information is missing?
    f. What is Dillard's average profit per day?
    g. On what day was the total value (in $) of returned goods the greatest? On what day was the total number of individual returned items the greatest?
    h. What is the maximum price paid for an item in our database? What is the minimum price paid for an item in our database?
    i. How many departments have more than 100 brands associated with them, and what are their descriptions?
    j. Retrieve the department descriptions of each of the skus in the skstinfo table.
    k. What department (with department description), brand, style, and color had the greatest total value of returned items?
    l. In what state and zip code is the store that had the greatest total revenue during the time period monitored in our dataset?
2. What metric should we use to assess sales trends?

a. How many distinct dates are there in the saledate column of the transaction table for each month/year combination in the database?

```
select distinct extract(year from saledate) as year_num,
extract(month from saledate) as month_num, saledate
from trnsact
group by year_num, month_num, saledate
order by year_num, month_num, saledate;
```

b. Determine which sku had the greatest total sales during the combined summer months of June, July, and August.

```
select sku,
      SUM(CASE WHEN extract(month from saledate)=6 THEN amt
ELSE 0 END) as JuneSales,
      SUM(CASE WHEN extract(month from saledate)=7 THEN amt
ELSE 0 END) as JulySales,
      SUM(CASE WHEN extract(month from saledate)=8 THEN amt
ELSE 0 END) as AugSales,
      JuneSales+JulySales+AugSales as TotalSales
from trnsact
where stype='P'
group by sku
order by TotalSales desc
```

c. How many distinct dates are there in the saledate column of the transaction table for each month/year/store combination in the database? Sort your results by the number of days per combination in ascending order.

```
select store, extract(year from saledate) as year_num, extract(month
from saledate) as month_num, count(distinct saledate) DistinctDates
from trnsact
group by store, year_num, month_num
order by DistinctDates asc;
```

d. What is the average daily revenue for each store/month/year combination in the database? Calculate this by dividing the total revenue for a group by the number of sales days available in the transaction table for that group.

```
select store, extract(year from saledate) as year_num, extract(month
from saledate) as month_num, count(distinct saledate)
NumTransactDays, SUM(amt)/NumTransactDays as AvgRevenue
from trnsact
where stype='P'
group by store, year_num, month_num
order by store, year_num, month_num asc;
```

3. How do the population statistics of the geographical location surrounding a store relate to sales performance?

    a. What is the average daily revenue brought in by Dillard's stores in areas of high, medium, or low levels of high school education?

```
select Max(Q.EduLevel),
sum(Q.TotalRevenue)/sum(Q.NumTransactDays) AvgDailyRev
from
(select
    t.store,
    extract(year from saledate) as year_num,
    extract(month from saledate) as month_num,
    count(distinct saledate) NumTransactDays,
    SUM(amt) as TotalRevenue,
    SUM(amt)/NumTransactDays as AvgRevenue,
    MAX(msa_high) HSGradRate,
    case
      when max(msa_high)>70 then 'HIGH'
      when max(msa_high)>60 and max(msa_high)<=70 then
'MEDIUM'
      when max(msa_high)>50 and max(msa_high)<=60 then 'LOW'
    end as EduLevel

from trnsact t join store_msa s on t.store=s.store
where stype='P' and extract(year from saledate)<>2005
group by t.store, year_num, month_num
having NumTransactDays>20) Q
group by Q.EduLevel
order by AvgDailyRev;
```

    b. Compare the average daily revenues of the stores with the highest median msa_income and the lowest median msa_income. In what

city and state were these stores, and which store had a higher average daily revenue?

```
select max(q.storenum), max(a.city), max(a.state),
max(q.incomelevel) income,
sum(q.totaldailyrevenue)/sum(q.numtransactdays) AvgDailyRevenue
from
(select
    t.store StoreNum,
    extract(month from saledate) as month_num,
    count(distinct saledate) NumTransactDays,
    SUM(amt) as TotalDailyRevenue,
    max(msa_income) as IncomeLevel

from trnsact t join store_msa s on t.store=s.store
where stype='P' and extract(year from saledate)<>2005
group by t.store, month_num
having NumTransactDays>20) q join strinfo a on q.storenum=a.store

where q.IncomeLevel = (select max(msa_income) from store_msa) or
q.IncomeLevel = (select min(msa_income) from store_msa)
group by q.storenum,q.incomelevel
```

c. What is the brand of the sku with the greatest standard deviation in sprice? Only examine skus that have been part of over 100 transactions.

```
select Q.item, stddev_samp(Q.saleprice) ItemStdDev, max(s.brand)
from
(select sku item, sprice saleprice
from trnsact
where stype='P'
group by sku, sprice
having count(distinct trannum)>100) as Q join skuinfo s on
Q.item=s.sku

group by q.item
order by itemstddev desc
```

d. Examine all the transactions for the sku with the greatest standard deviation in sprice, but only consider skus that are part of more than 100 transactions.

```
select extract(month from saledate) as month_num,
SUM(amt)/count(distinct saledate) as AvgDailyRevenue
 from trnsact
 where stype='P' and extract(year from saledate)<>2005
 group by month_num
having count(distinct saledate)>20
order by AvgDailyRevenue desc;
```

4. Analyzing monthly (or seasonal) sales effects

a. What was the average daily revenue Dillard's brought in during each month of the year?

```
select extract(month from saledate) as month_num,
SUM(amt)/count(distinct saledate) as AvgDailyRevenue
 from trnsact
 where stype='P' and extract(year from saledate)<>2005
 group by month_num
having count(distinct saledate)>20
order by AvgDailyRevenue desc;
```

b. Which department, in which city and state of what store, had the greatest % increase in average daily sales revenue from November to December?

```
select q.storename, Q.item, Q.deptname, R.city, R.state,
Q.PercentSalesIncrease
from
(
select
    t.store storename,
    t.sku item,
    max(s.dept) deptnum,
    max(d.deptdesc) deptname,
    COUNT(DISTINCT (CASE WHEN extract(month from saledate)=11
THEN saledate END)) as NovSaleDates,
    COUNT(DISTINCT (CASE WHEN extract(month from saledate)=12
THEN saledate END)) as DecSaleDates,
```

```
    SUM(CASE WHEN extract(month from saledate)=11 THEN amt
ELSE 0 END) as TotNovSales,
    SUM(CASE WHEN extract(month from saledate)=12 THEN amt
ELSE 0 END) as TotDecSales,
    CASE WHEN NovSaleDates<>0 THEN TotNovSales/NovSaleDates
ELSE TotNovSales END as NovAvgDailySales,
    CASE WHEN DecSaleDates<>0 THEN TotDecSales/DecSaleDates
ELSE TotDecSales END as DecAvgDailySales,
    CASE WHEN NovAvgDailySales<>0 and
DecAvgDailySales>NovAvgDailySales THEN (DecAvgDailySales-
NovAvgDailySales)*100/NovAvgDailySales ELSE 0 END as
PercentSalesIncrease

from trnsact t join skuinfo s on t.sku=s.sku join deptinfo d on
s.dept=d.dept
where t.stype='P' and extract(year from t.saledate)<>2005
group by t.store, t.sku
having count(distinct saledate)>20) as Q

join strinfo R on Q.storename=R.store
order by Q.PercentSalesIncrease desc;
```

c. What is the city and state of the store that had the greatest decrease in average daily revenue from August to September?

```
select
    t.store storename,
    t.sku item,
    COUNT(DISTINCT (CASE WHEN extract(month from saledate)=8
THEN saledate END)) as AugSaleDates,
    COUNT(DISTINCT (CASE WHEN extract(month from saledate)=9
THEN saledate END)) as SepSaleDates,
    SUM(CASE WHEN extract(month from saledate)=8 THEN amt ELSE
0 END) as TotAugSales,
    SUM(CASE WHEN extract(month from saledate)=9 THEN amt ELSE
0 END) as TotSepSales,
    CASE WHEN AugSaleDates<>0 THEN TotAugSales/AugSaleDates
ELSE TotAugSales END as AugAvgDailySales,
    CASE WHEN SepSaleDates<>0 THEN TotSepSales/SepSaleDates
ELSE TotSepSales END as SepAvgDailySales,
```

```
    CASE WHEN SepAvgDailySales<AugAvgDailySales THEN
AugAvgDailySales-SepAvgDailySales ELSE 0 END as SalesDecrease

from trnsact t join skuinfo s on t.sku=s.sku
where t.stype='P' and extract(year from t.saledate)<>2005
group by t.store, t.sku
having count(distinct saledate)>20) as Q

join strinfo R on Q.storename=R.store
order by Q.SalesDecrease desc;
```

d. Determine the month of maximum total revenue for each store. Count the number of stores whose month of maximum total revenue was in each of the twelve months. Then determine the month of maximum average daily revenue. Count the number of stores whose month of maximum average daily revenue was in each of the twelve months. How do they compare?

```
select
    case
      when Month_Num=8 THEN 'AUG'
      when Month_Num=9 THEN 'SEP'
      when Month_Num=10 THEN 'OCT'
      when Month_Num=11 THEN 'NOV'
      when Month_Num=12 THEN 'DEC'
    end,
    count(Q.Store_Num) as Num_Stores
from
(
select
   store Store_Num,
   extract(month from saledate) as Month_num,
   count(distinct saledate) as NumTransactDays,
   SUM(amt) as TotalDailyRevenue,
   SUM(amt)/NumTransactDays as AvgDailyRevenue,
   ROW_NUMBER() OVER (PARTITION BY Store_Num ORDER BY
AvgDailyRevenue asc) as SalesRank
 from trnsact
 where stype='P' and extract(year from saledate)<>2005
 group by store, month_num
having NumTransactDays>20
```

```
qualify SalesRank=1
) as Q
group by Q.Month_Num
order by Month_Num
```