Applications Needed:

[\\corpfs-v\Software\Development\Visual Studio 2017\vs\_enterprise\_\_397464819.1516976509.exe](file:///\\corpfs-v\Software\Development\Visual%20Studio%202017\vs_enterprise__397464819.1516976509.exe)

[\\corpfs-v\Software\Development\Visual Studio 2017\LicenseKey.txt](file:///\\corpfs-v\Software\Development\Visual%20Studio%202017\LicenseKey.txt)

[\\corpfs-v\Software\Development\SSDT 17.1\SSDTSetup.exe](file:///\\corpfs-v\Software\Development\SSDT%2017.1\SSDTSetup.exe)

[\\corpfs-v\Software\Development\SQL2017\SSMS\SSMS-Setup-ENU.exe](file:///\\corpfs-v\Software\Development\SQL2017\SSMS\SSMS-Setup-ENU.exe)

<https://cran.rstudio.com/>

<https://www.rstudio.com/products/rstudio/download/>

<https://www.anaconda.com/download/>

**SQL**

**Servers:**

* Azure Connections:
  + Dev: [hy-vee-dw-sql-server-d.database.windows.net](http://hy-vee-dw-sql-server-d.database.windows.net/)
  + QA: [hy-vee-dw-sql-server-q.database.windows.net](http://hy-vee-dw-sql-server-q.database.windows.net/)
  + Prod: [hy-vee-dw-sql-server-p.database.windows.net](http://hy-vee-dw-sql-server-p.database.windows.net/)
    - Only use for production ready projects
* \*\*QA should be an exact copy of Prod, and better than what DR has been in the past. Dev will be up to date, but will have only samples of data for us to test on. We should test on dev first where possible to get queries running before running on QA\*\*
* Fuel, HR and WHS Marts (not in Azure):
* Prod: PWWDMDW-SQL1
* QA: QWWDMDW-SQL1
* Dev: DWWDMDW-SQL1
* **SSMS**
  + **Dev**
    - Devsrv11-v
    - Devsrv2\corp
    - Devsrv14-v
    - Hvcorp14qa-v
  + **Prod**
    - Hvcorp11
    - Hvcorp10\corp
    - Hvcorp14

DIM\_ODS.Nielsen\_MarketData is a week behind Hy-Vee’s sales data. For example, if the most recent week of sales data is through 11-10-19, then we will only have rest of market data through 11-03-19. Also, Nielsen Market Data runs from Sunday through Saturday, while the Hy-Vee’s POS data runs from Monday through Sunday, so adjustments need to be made when linking internal sales data with Nielsen data.

Use POS\_Department when

Use Corporate\_Department when

Item “Description” is not unique, and there is no central monitoring of this attribute (stores can update at will) so it should be avoided in trying to identify items. A UPC can be tied to multiple Descriptions, but a Description will only be tied to one UPC. Use UPC or UPC/Description combo to identify items.

**POS Tables:**

* SalesLineItemFact
  + What: Every basket and item for every customer. When a customer purchases 10 items, 10 rows are created
  + Where: hvcorp11, devsrv11-v, P\_POS\_MART, T\_POS\_MART
  + When to use: when you need both item and customer level information
* SalesDailyItemFact
  + What: Every item’s daily sales for every location. The customer grouping is taken out of this table
  + Where: hvcorp11, devsrv11-v, P\_POS\_MART, T\_POS\_MART
  + When to use: when you need just item level information, and you don’t need any customer info
* SalesBasketFact
  + What: Every customer’s basket. One purchase is one row
  + Where: P\_POS\_MART, T\_POS\_MART
  + When to use: when you are interested in baskets or customer info and don’t need items

Dimension Tables:

* StoreItemDimension
* LocationDimension
* CustomerDimension
* DateDimension
* POSDepartmentDimension

1. Find the subcategory with the greatest increase in sales between the second to last completed fiscal week and the last completed fiscal week. Always get the last two completed weeks
2. Get average sales and baskets per fiscal week by facts segment for the last completed fiscal month. Dynamic to always get the last full month
3. Find the customers that purchased more grocery department items than the average Hy-Vee shopper in the last 30 days
4. By store, show last month’s private label and online sales, and compare to what they were the year before. Display sales and baskets and percent change in both
5. Find the ad group that had the greatest lift in sales from the last completed mega ad compared to the 30 days prior to the ad
6. By shopstyle, what percent of food store customers had a purchase at a c-store, and vice versa? Also group by customers whose frequent store has a c-store and those whose frequent store does not. Use last 60 days
7. Pick a customer and identify what % of the items they have bought in the last 60 days were bought in at least 75% of their baskets, what % were bought in at least 50% of their baskets, what % were bought in at least 25% of their baskets, and what % were bought in less than 25% of their baskets
8. Take the info from problem 7 and get the data by facts and shopstyle segments
9. Get a list of customers that have purchased soft drinks but not pepsi in the last 30 days. Order by distinct items bought in the soft drink subcategory

**R/Python**

* Data: a week during recent sales spike and a week before (spike started week ending 9-23)
* Get a flag column for if they shopped prior to the week starting
* Compare data for customers who shopped both weeks, and the customers who shopped just one of the 2 weeks. How are they different? How are they the same? Look for a couple things you think might explain the difference and say whether or not they are different between the two groups, and by how much
* Look at the gained customers from the second week and see if they are different from all customers in week one