# COMPLETED
# SAMPLE OF WORK
# DATA SCIENCE
# (Employee Churn in R)

NORIE JEANNE PEREIRA

Steps:

1. Import dataset()MFG10YearTerminationData.csv
   (First Look at The Data- The Structure)
   # Load an R data frame:

   # Load an R data frame:
   MFG10YearTerminationData <- read.csv("~/5 Data Science Project Tuturial/HR Analytics in R & Python/Prime Manpower HR Analytics/csv/MFG10YearTerminationData.csv")
   MYdataset <- MFG10YearTerminationData
   str(MYdataset

2. Read csv

```
   'data.frame':   49653 obs. of  18 variables:
 $ EmployeeID         : int  1318 1318 1318 1318 1318 1318 1318 1318 1318 1318 ...
 $ recorddate_key     : Factor w/ 130 levels "1/1/2006 0:00",..: 41 42 43 44 45 46
47 48 49 50 ...
 $ birthdate_key      : Factor w/ 5342 levels "1941-01-15","1941-02-14",..: 1075
1075 1075 1075 1075 1075 1075 1075 1075 1075 ...
 $ orighiredate_key   : Factor w/ 4415 levels "1989-08-28","1989-08-31",..: 1 1 1
1 1 1 1 1 1 1 ...
 $ terminationdate_key: Factor w/ 1055 levels "1900-01-01","2006-01-01",..: 1 1 1
1 1 1 1 1 1 1 ...
 $ age                : int  52 53 54 55 56 57 58 59 60 61 ...
 $ length_of_service  : int  17 18 19 20 21 22 23 24 25 26 ...
 $ city_name          : Factor w/ 40 levels "Abbotsford","Aldergrove",..: 35 35 35
35 35 35 35 35 35 35 ...
 $ department_name    : Factor w/ 21 levels "Accounting","Accounts Payable",..: 10
10 10 10 10 10 10 10 10 10 ...
 $ job_title          : Factor w/ 47 levels "Accounting Clerk",..: 9 9 9 9 9 9 9 9
9 9 ...
 $ store_name         : int  35 35 35 35 35 35 35 35 35 35 ...
 $ gender_short       : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ gender_full        : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2
2 ...
 $ termreason_desc    : Factor w/ 4 levels "Layoff","Not Applicable",..: 2 2 2 2 2
2 2 2 2 2 ...
 $ termtype_desc      : Factor w/ 3 levels "Involuntary",..: 2 2 2 2 2 2 2 2 2 2
2 ...
 $ STATUS_YEAR        : int  2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 ...
 $ STATUS             : Factor w/ 2 levels "ACTIVE","TERMINATED": 1 1 1 1 1 1 1 1
1 1 ...
 $ BUSINESS_UNIT      : Factor w/ 2 levels "HEADOFFICE","STORES": 1 1 1 1 1 1 1 1
1 1 ...
```

   3.

4. Second Look at The Data- Data Quality

```
      EmployeeID           recorddate_key     birthdate_key     orighiredate_key
 Min.   :1318   12/31/2013 0:00: 5215   1954-08-04:   40   1992-08-09:   50
 1st Qu.:3360   12/31/2012 0:00: 5101   1956-04-27:   40   1995-02-22:   50
 Median :5031   12/31/2011 0:00: 4972   1973-03-23:   40   2004-12-04:   50
 Mean   :4859   12/31/2014 0:00: 4962   1952-01-27:   30   2005-10-16:   50
 3rd Qu.:6335   12/31/2010 0:00: 4840   1952-08-10:   30   2006-02-26:   50
 Max.   :8336   12/31/2015 0:00: 4799   1953-10-06:   30   2006-09-25:   50
                (Other)        :19764   (Other)   :49443   (Other)   :49353
 terminationdate_key      age        length_of_service
 1900-01-01:42450    Min.   :19.00   Min.   : 0.00
 2014-12-30: 1079    1st Qu.:31.00   1st Qu.: 5.00
```

```
 2015-12-30:  674    Median :42.00   Median :10.00
 2010-12-30:   25    Mean   :42.08   Mean   :10.43
 2012-11-11:   21    3rd Qu.:53.00   3rd Qu.:15.00
 2015-02-04:   20    Max.   :65.00   Max.   :26.00
 (Other)   : 5384
          city_name             department_name          job_title
 Vancouver      :11211   Meats           :10269   Meat Cutter  :9984
 Victoria       : 4885   Dairy           : 8599   Dairy Person :8590
 Nanaimo        : 3876   Produce         : 8515   Produce Clerk:8237
 New Westminster: 3211   Bakery          : 8381   Baker        :8096
 Kelowna        : 2513   Customer Service: 7122   Cashier      :6816
 Burnaby        : 2067   Processed Foods : 5911   Shelf Stocker:5622
 (Other)        :21890   (Other)         :  856   (Other)      :2308
store_name   gender_short gender_full          termreason_desc
 Min.   : 1.0   F:25898      Female:25898   Layoff        : 1705
 1st Qu.:16.0   M:23755      Male  :23755   Not Applicable:41853
 Median :28.0                               Resignaton    : 2111
 Mean   :27.3                               Retirement    : 3984
 3rd Qu.:42.0
 Max.   :46.0

  termtype_desc    STATUS_YEAR          STATUS         BUSINESS_UNIT
 Involuntary   : 1705   Min.   :2006   ACTIVE    :48168   HEADOFFICE:  585
 Not Applicable:41853   1st Qu.:2008   TERMINATED: 1485   STORES    :49068
 Voluntary     : 6095   Median :2011
                        Mean   :2011
                        3rd Qu.:2013
                        Max.   :2015
```

5. Third Look at the Data – Generally What Is The Data Telling Us?
   What proportion of our staff are leaving?

| | EmployeeID <int> | recorddate_key <chr> | birthdate_key <chr> | orighiredate_key <chr> | terminationdate_key <chr> | age <int> | length_of_service <int> | city_name <chr> |
|---|---|---|---|---|---|---|---|---|
| 1 | 1318 | 12/31/2006 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 52 | 17 | Vancouver |
| 2 | 1318 | 12/31/2007 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 53 | 18 | Vancouver |
| 3 | 1318 | 12/31/2008 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 54 | 19 | Vancouver |
| 4 | 1318 | 12/31/2009 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 55 | 20 | Vancouver |
| 5 | 1318 | 12/31/2010 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 56 | 21 | Vancouver |
| 6 | 1318 | 12/31/2011 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 57 | 22 | Vancouver |
| 7 | 1318 | 12/31/2012 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 58 | 23 | Vancouver |

| | EmployeeID <int> | recorddate_key <chr> | birthdate_key <chr> | orighiredate_key <chr> | terminationdate_key <chr> | age <int> | length_of_service <int> | city_name <chr> |
|---|---|---|---|---|---|---|---|---|
| 8 | 1318 | 12/31/2013 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 59 | 24 | Vancouver |
| 9 | 1318 | 12/31/2014 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 60 | 25 | Vancouver |
| 10 | 1318 | 12/31/2015 0:00 | 1954-01-03 | 1989-08-28 | 1900-01-01 | 61 | 26 | Vancouver |
| 11 | 1319 | 12/31/2006 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 49 | 17 | Vancouver |
| 12 | 1319 | 12/31/2007 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 50 | 18 | Vancouver |
| 13 | 1319 | 12/31/2008 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 51 | 19 | Vancouver |
| 14 | 1319 | 12/31/2009 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 52 | 20 | Vancouver |
| 15 | 1319 | 12/31/2010 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 53 | 21 | Vancouver |
| 16 | 1319 | 12/31/2011 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 54 | 22 | Vancouver |
| 17 | 1319 | 12/31/2012 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 55 | 23 | Vancouver |
| 18 | 1319 | 12/31/2013 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 56 | 24 | Vancouver |
| 19 | 1319 | 12/31/2014 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 57 | 25 | Vancouver |
| 20 | 1319 | 12/31/2015 0:00 | 1957-01-03 | 1989-08-28 | 1900-01-01 | 58 | 26 | Vancouver |
| 21 | 1320 | 12/31/2006 0:00 | 1955-01-02 | 1989-08-28 | 1900-01-01 | 51 | 17 | Vancouver |

*Next*

*123456*

*...*

*48*

*Previous*

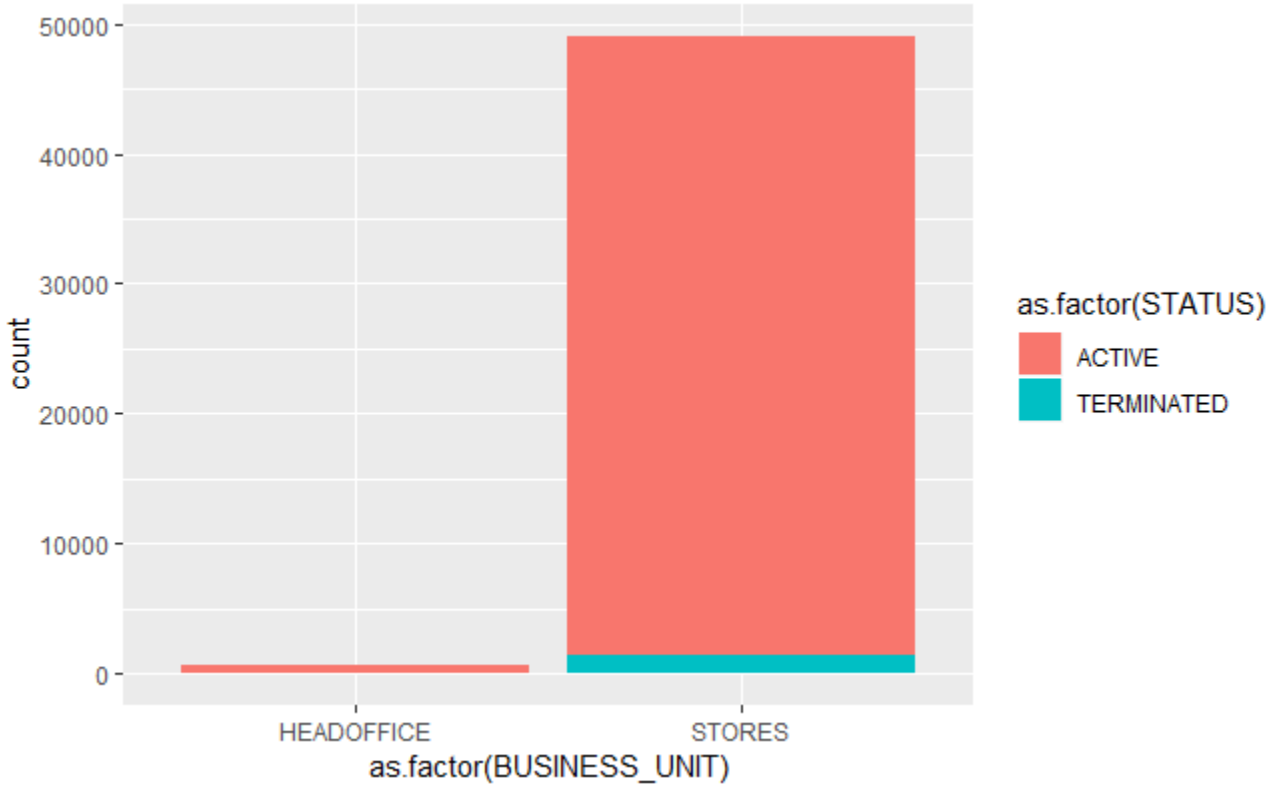
*1-21 of 49,653 rows | 1-9 of 18 columns*

6. mean(StatusCount$PercentTerminated)

```
2.997124
```

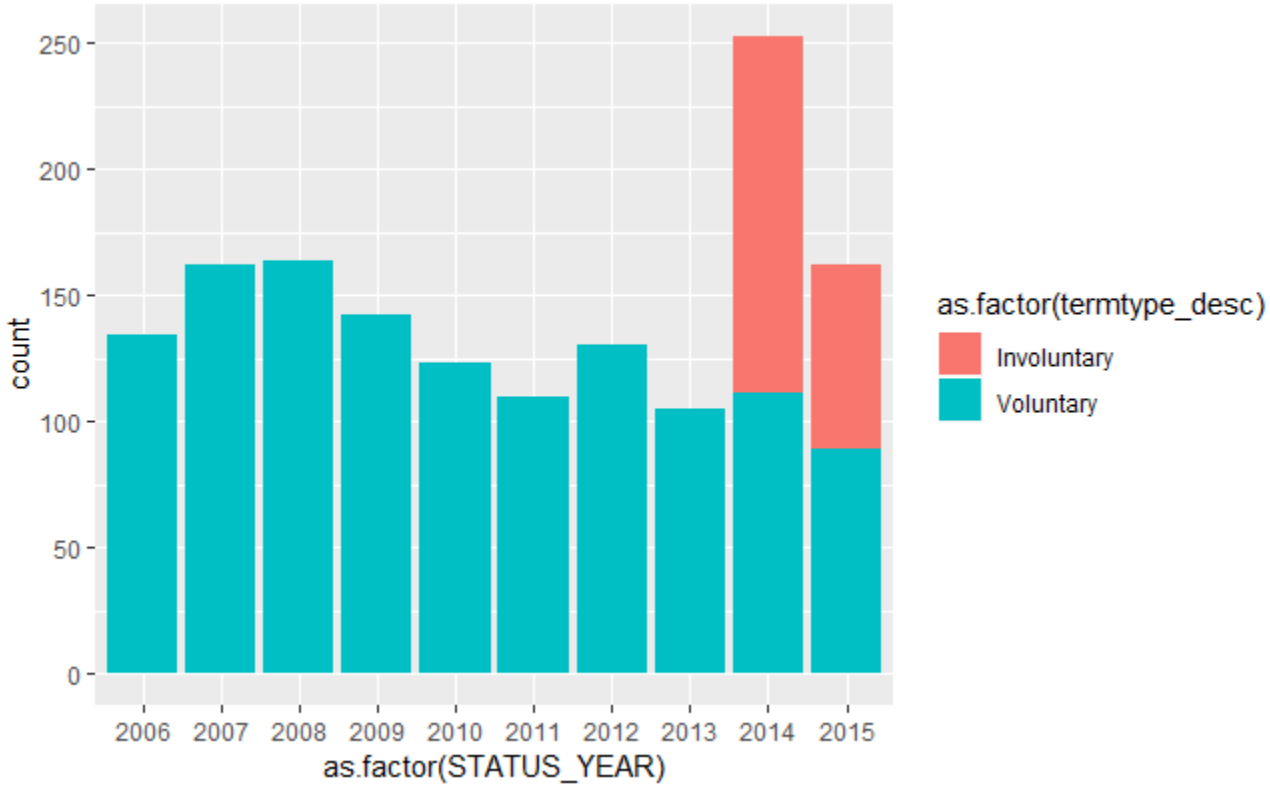
It looks like it ranges from 1.97 to 4.85% with an average of 2.99%

7. Where are the terminations occurring?
   Chart-By Business Unit

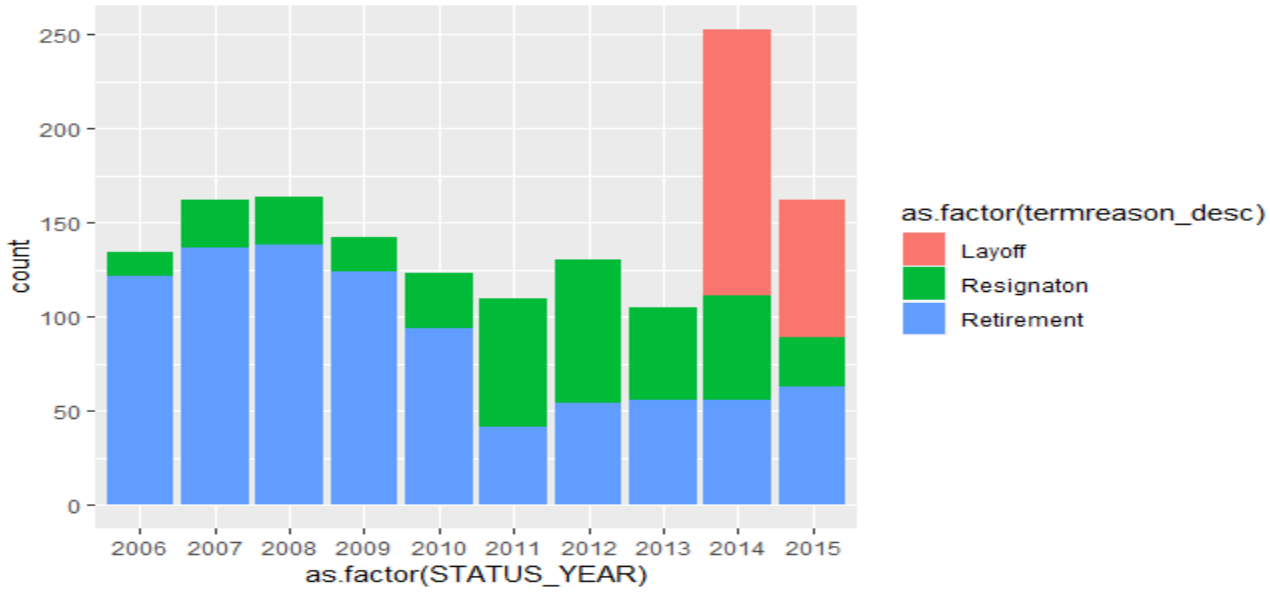



Termination is the last 10 years have predominantly occurred in the STORES business unit. Only 1 termination in HR Technology which is in the head office.

8. Termination Type And Status Year

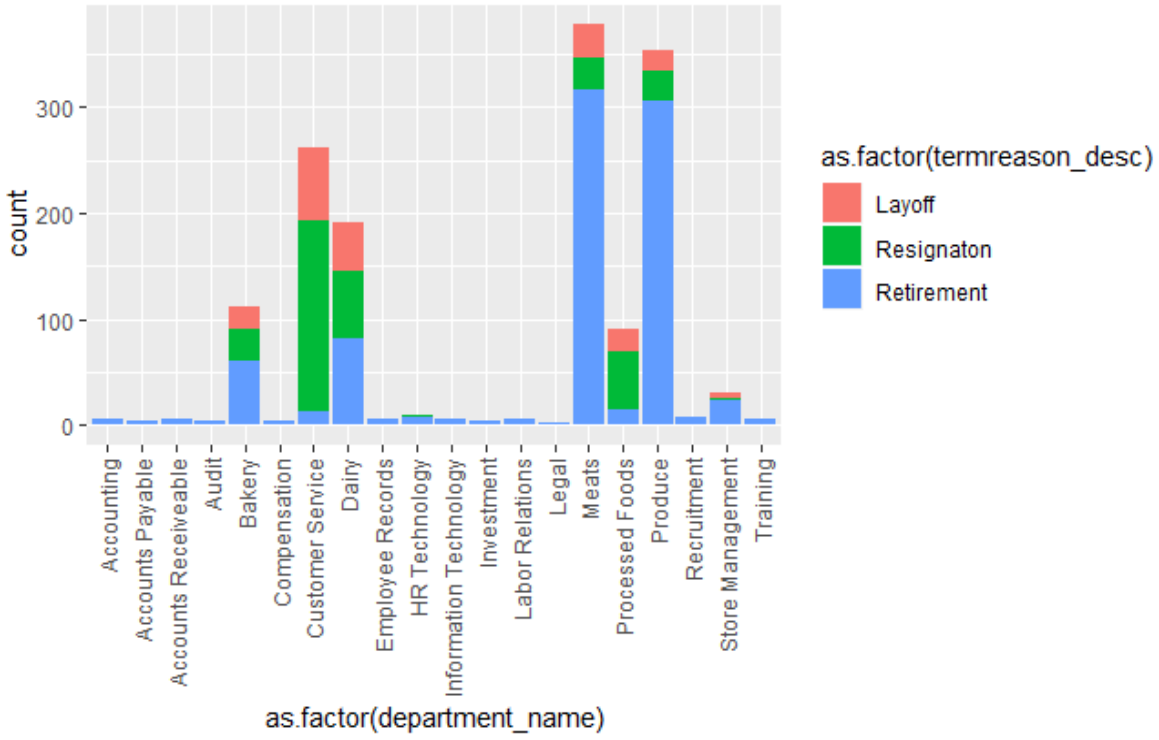



9. By Status Year and Termination Reason




It seems that there were layoffs in 2014 and 2015 which accounts for the involuntary terminations.

10. Terminates By Termination Reason and Department




When we look at the terminate by Department, Customer Service has a much larger proportion of resignation compared to other departments. And retirement in general is high is a number of departments.