Project2 Data 607

Umais Siddiqui

October 7, 2017

Rpubs Link: http://rpubs.com/umais/data607_Project2

Github Link: https://github.com/umais/DATA-607/tree/master/Project2_Data607

Down Stream Analysis using tidyR and Dplyer

I will be using three data sets from a past projects that I worked on. One of the data sets is from the company I work for at the moment.

School Data Set

1) Data set from a schools project I did for couple of schools back home. The data shows the number of students that passed and failed from 9th and 10th grade . I have deindentified the data by using fake school names

```
mydb = dbConnect(MySQL(), user='root', password='Welcome@1', dbname='project2', host='localhost')

rs = dbSendQuery(mydb, "SELECT * FROM StudentRecords;")

df=fetch(rs, n=-1)

head(df)
```

```
##   School_Id SchoolName NinthGradePass NinthGradeFail TenthGradePass
## 1         1    School1            189             96            145
## 2         2    School2            167             34            145
## 3         3    School3            178             34            187
##   TenthGradeFail
## 1             23
## 2             12
## 3              3
```

Summarise the Data in to total Passed and Failed from each school

```
#Gather Function from tidyR
TotalStudents<- gather(df,Status,NumberOfStudents,NinthGradePass:TenthGradeFail)

head(TotalStudents)
```

```
##   School_Id SchoolName         Status NumberOfStudents
## 1         1    School1 NinthGradePass              189
## 2         2    School2 NinthGradePass              167
## 3         3    School3 NinthGradePass              178
## 4         1    School1 NinthGradeFail               96
## 5         2    School2 NinthGradeFail               34
```

```
## 6         3    School3 NinthGradeFail               34
```

```
failed=TotalStudents %>% dplyr::filter(Status %like% "Fail")%>%group_by(SchoolName)%>%summarise(total=su

failed
```

```
## # A tibble: 3 × 2
##   SchoolName total
##        <chr> <int>
## 1    School1   119
## 2    School2    46
## 3    School3    37
```

```
passed=TotalStudents %>% dplyr::filter(Status %like% "Pass")%>%group_by(SchoolName)%>%summarise(total=su

passed
```

```
## # A tibble: 3 × 2
##   SchoolName total
##        <chr> <int>
## 1    School1   334
## 2    School2   312
## 3    School3   365
```

Second Data Set for In Network and Out of Network Providers

```
rs = dbSendQuery(mydb, "SELECT * FROM ProviderNetwork ;")

df=fetch(rs, n=-1)

head(df)
```

```
##   Payer_Id          PayerName TotalInNetworkChicago
## 1        1 Health Insurance A                   445
## 2        2 Health Insurance B                   335
##   TotalOutOfNetworkChicago TotalInNetwork_NewYork
## 1                       45                    665
## 2                       23                    667
##   TotalOutOfNetwork_NewYork
## 1                       234
## 2                       230
```

```
TotalProviders<- gather(df,NetworkStatus,NumberOfProviders,TotalInNetworkChicago:TotalOutOfNetwork_NewY

TotalProviders
```

```
##   Payer_Id          PayerName             NetworkStatus NumberOfProviders
## 1        1 Health Insurance A     TotalInNetworkChicago               445
## 2        2 Health Insurance B     TotalInNetworkChicago               335
## 3        1 Health Insurance A  TotalOutOfNetworkChicago                45
## 4        2 Health Insurance B  TotalOutOfNetworkChicago                23
## 5        1 Health Insurance A    TotalInNetwork_NewYork               665
## 6        2 Health Insurance B    TotalInNetwork_NewYork               667
## 7        1 Health Insurance A TotalOutOfNetwork_NewYork               234
## 8        2 Health Insurance B TotalOutOfNetwork_NewYork               230
```

Check Which Health Plan has More In Network Providers from the two cities.

```
InNetwork=TotalProviders %>% dplyr::filter(NetworkStatus %like% "InNetwork")%>%group_by(PayerName)%>%su

InNetwork
```

```
## # A tibble: 2 × 2
##            PayerName total
##                <chr> <int>
## 1 Health Insurance A  1110
## 2 Health Insurance B  1002
```

As we can See Payer 1 has more in network providers than payer 2.

Using the Third data set for the MemberRisk Data

```
rs = dbSendQuery(mydb, "SELECT * FROM  MemberRisk ;")

df=fetch(rs, n=-1)

head(df)
```

```
##   Organization_Id OrganizationName HighRiskFlorida HighRiskNewYork
## 1               1        AlignCare             550             660
## 2               2  Demo Data Large             345             660
##   LowRiskFlorida LowRiskNewYork
## 1            220            330
## 2            286            390
```

```
TotalMembers<- gather(df,RiskStatus,NumberOfMembers,HighRiskFlorida:LowRiskNewYork)

TotalMembers
```

```
##   Organization_Id OrganizationName      RiskStatus NumberOfMembers
## 1               1        AlignCare HighRiskFlorida             550
## 2               2  Demo Data Large HighRiskFlorida             345
## 3               1        AlignCare HighRiskNewYork             660
## 4               2  Demo Data Large HighRiskNewYork             660
## 5               1        AlignCare  LowRiskFlorida             220
## 6               2  Demo Data Large  LowRiskFlorida             286
## 7               1        AlignCare  LowRiskNewYork             330
## 8               2  Demo Data Large  LowRiskNewYork             390
```

```
HighRisk=TotalMembers %>% dplyr::filter(RiskStatus %like% "High")%>%group_by(OrganizationName)%>%summar

HighRisk
```

```
## # A tibble: 2 × 2
##   OrganizationName total
##              <chr> <int>
## 1        AlignCare  1210
## 2  Demo Data Large  1005
```

In the last Data Set we can see that AlignCare has more High Risk Members than Demo Data Large