Mistplay Data Analysis

Aymen Rumi

10/10/2020

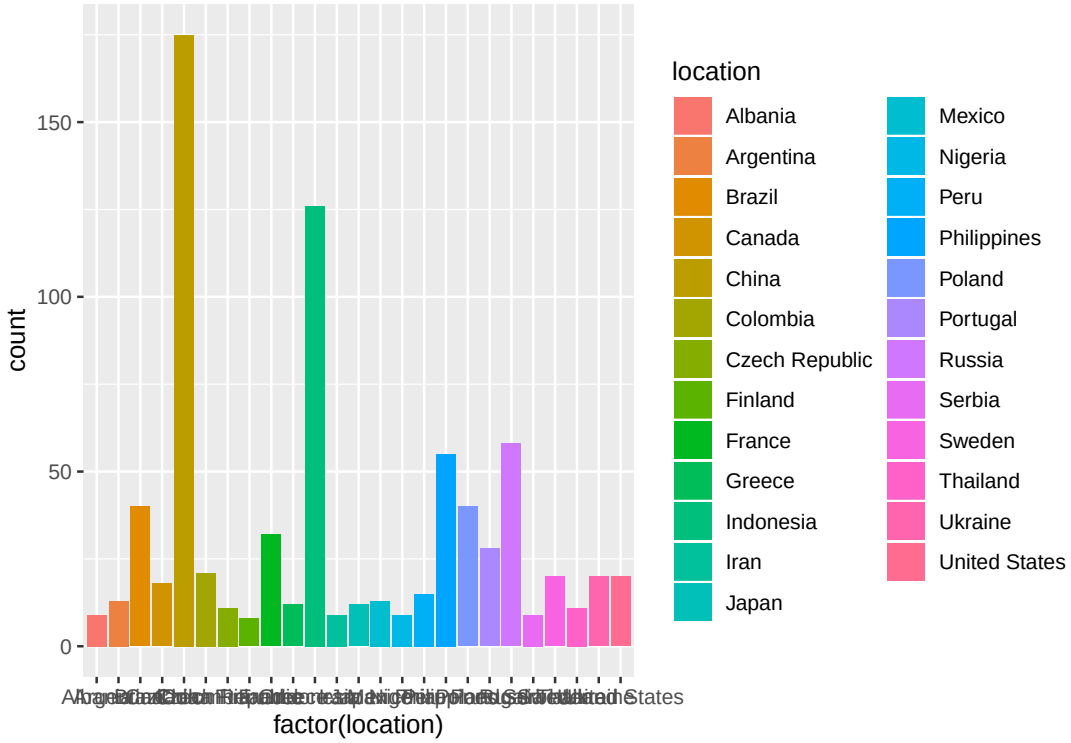
Data Analysis for Mistplay Data Engineering Dataset

Data Visuzalization & Summary Statistic for Dataset given for Mistplay Data Engineering Challenge

1. Distribution of Location
2. Distribution of Age Group
3. Total Revenue by Location
4. Total Revenue by Age Group
5. Total Revenue by Location per Person

Distribution for Location -Top 25

```
location_distribution<-data%>%group_by(location)%>%summarise(count=n())%>%mutate(prop=count/sum(count))%

ggplot(location_distribution, aes(factor(location), count, fill = location)) +
    geom_col(position = 'dodge')
```

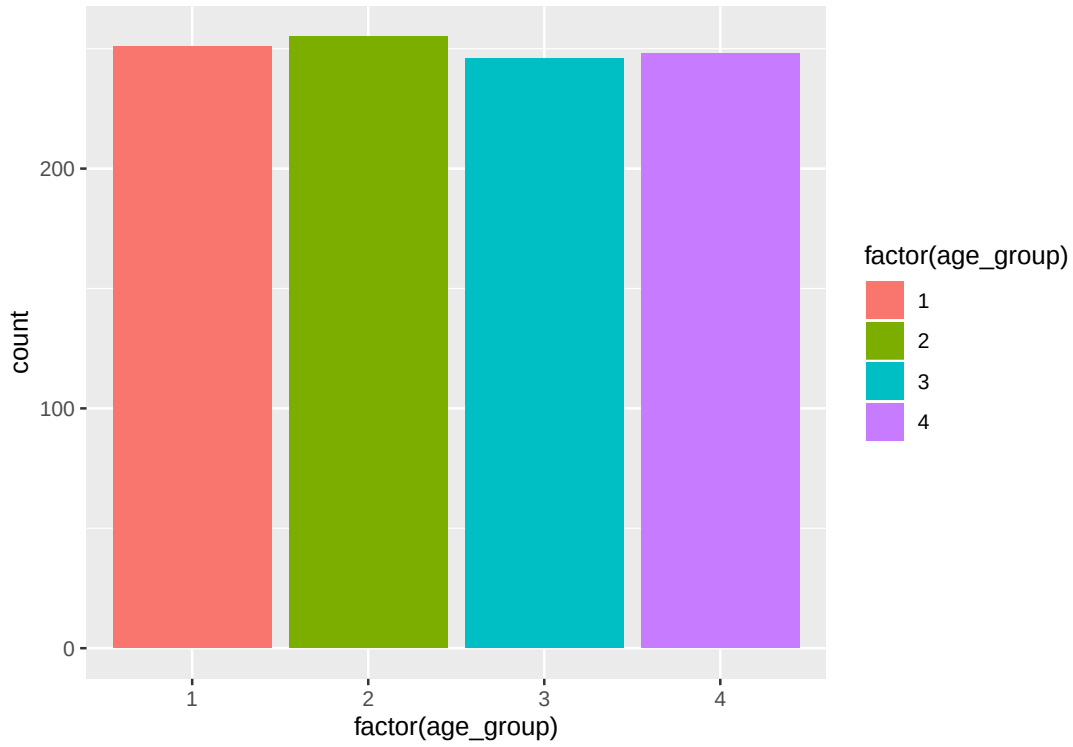


```
location_distribution%>%kable()
```

location	count	prop
China	175	0.175
Indonesia	126	0.126
Russia	58	0.058
Philippines	55	0.055
Brazil	40	0.040
Poland	40	0.040
France	32	0.032
Portugal	28	0.028
Colombia	21	0.021
Sweden	20	0.020
Ukraine	20	0.020
United States	20	0.020
Canada	18	0.018
Peru	15	0.015
Argentina	13	0.013
Mexico	13	0.013
Greece	12	0.012
Japan	12	0.012
Czech Republic	11	0.011
Thailand	11	0.011
Albania	9	0.009
Iran	9	0.009
Nigeria	9	0.009
Serbia	9	0.009
Finland	8	0.008

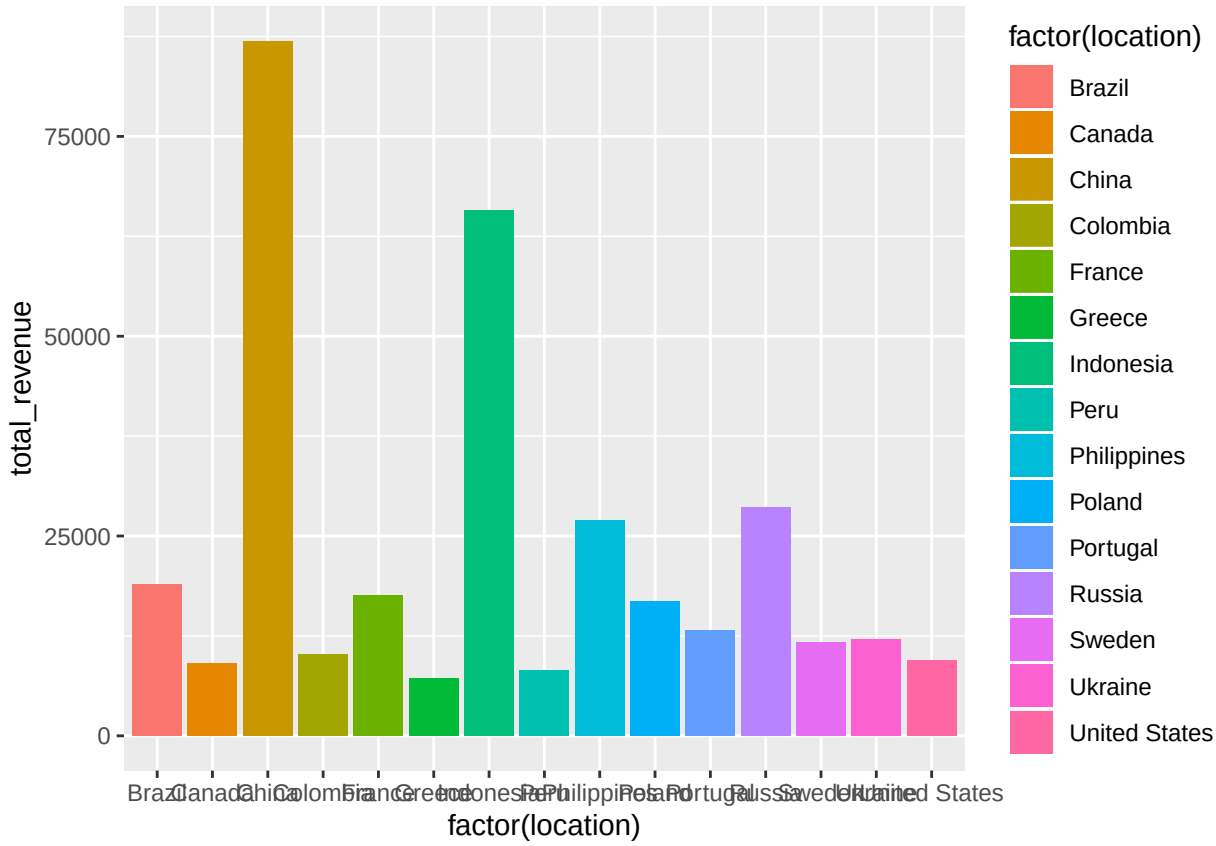
Distribution for Age Group

```
age_distribution<-data%>%group_by(age_group)%>%summarise(count=n())%>%mutate(prop=count/sum(count))%>%a

ggplot(age_distribution, aes(factor(age_group), count, fill = factor(age_group))) +
    geom_col(position = 'dodge')
```



Total Revenue by Location - Top 15

```
revenue_by_location<-data%>%group_by(location)%>%summarise(total_revenue=sum(revenue))%>%arrange(desc(t

ggplot(revenue_by_location, aes(factor(location), total_revenue, fill = factor(location))) +
    geom_col(position = 'dodge')
```



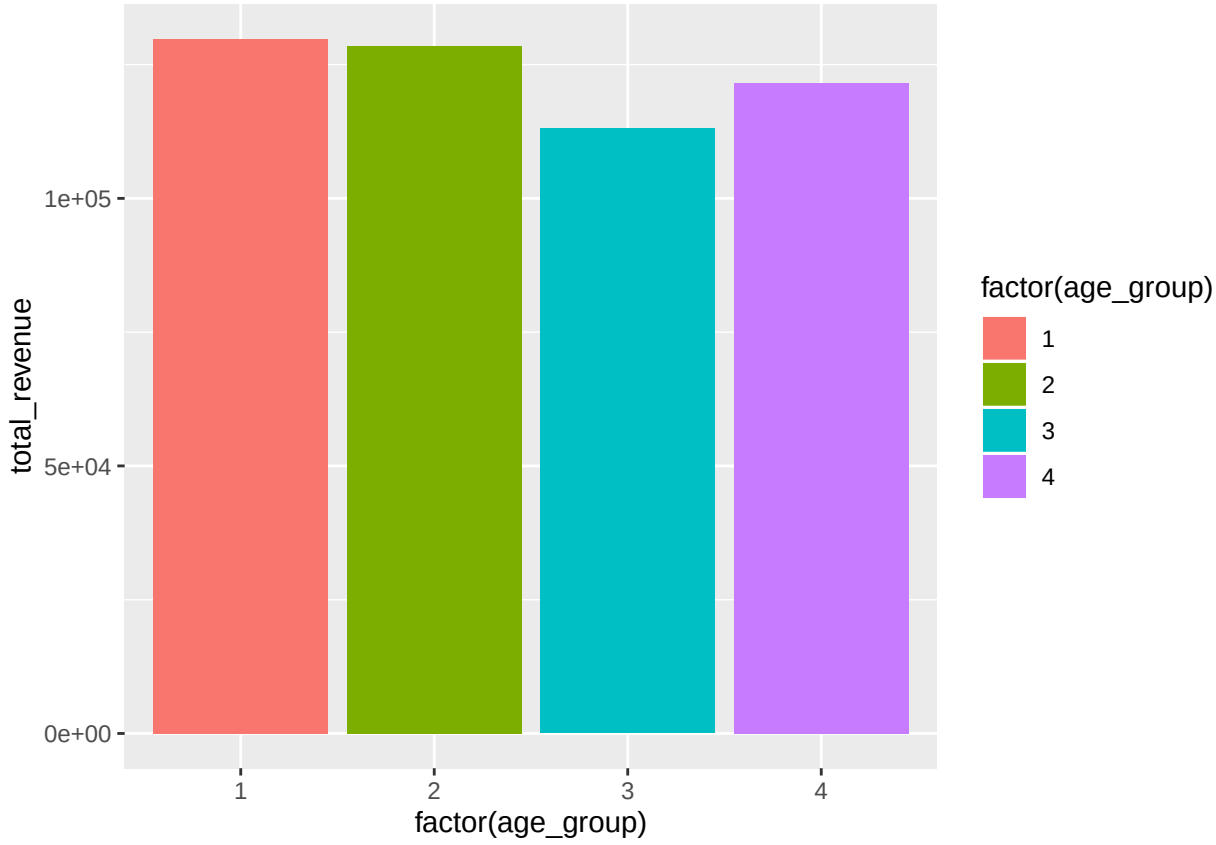

```
revenue_by_location%>%kable()
```

location	total_revenue
China	86934.94
Indonesia	65801.21
Russia	28640.84
Philippines	26957.02
Brazil	18978.00
France	17534.37
Poland	16799.83
Portugal	13215.37
Ukraine	12099.07
Sweden	11651.32
Colombia	10220.70
United States	9427.31
Canada	9033.20
Peru	8189.84
Greece	7227.59

Total Revenue by Age Group

```
revenue_by_age<-data%>%group_by(age_group)%>%summarise(total_revenue=sum(revenue))%>%arrange(desc(total_
```

```
ggplot(revenue_by_age, aes(factor(age_group), total_revenue, fill = factor(age_group))) +
    geom_col(position = 'dodge')
```

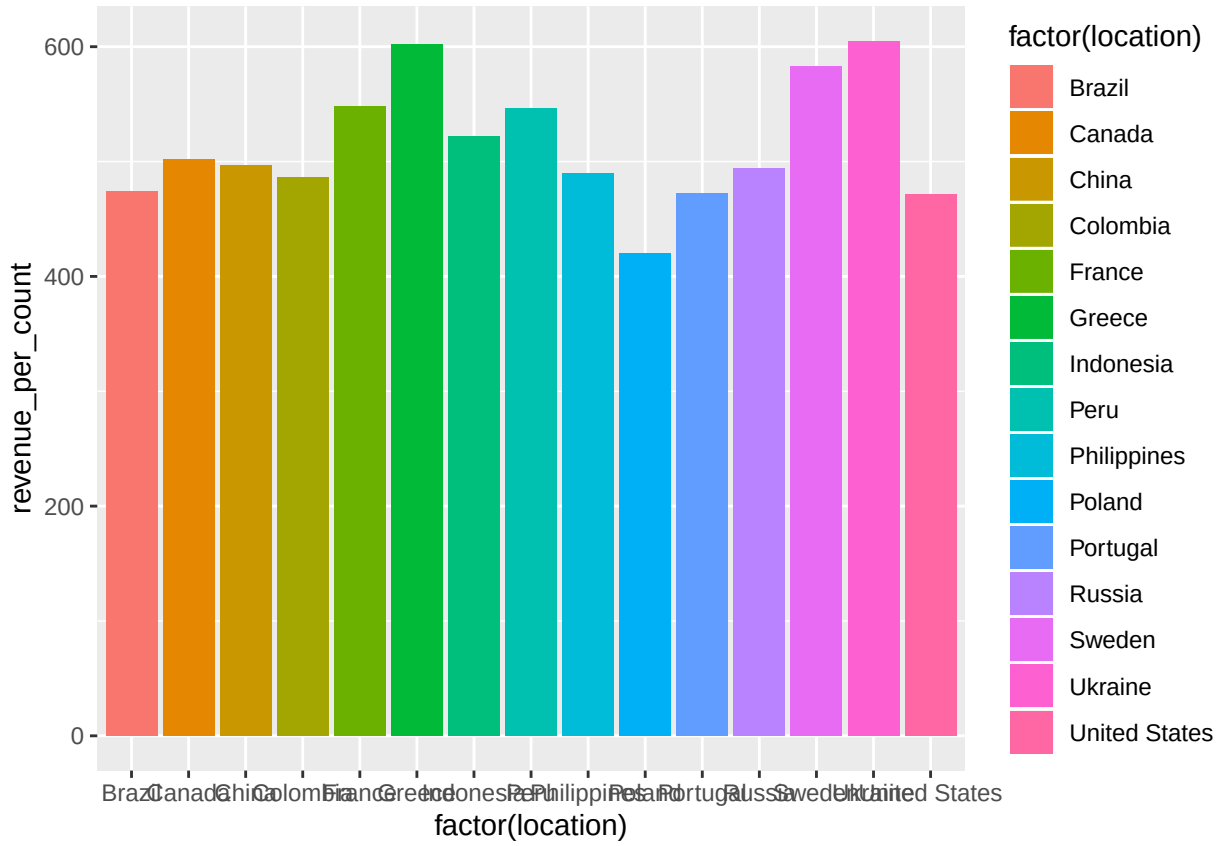



```
revenue_by_age%>%kable()
```

age_group	total_revenue
1	129846.5
2	128499.9
4	121559.7
3	113059.1

Total Revenue by Location per Person

```
revenue_by_location_per_person<-data%>%group_by(location)%>%summarise(total_revenue=sum(revenue),count=n

ggplot(revenue_by_location_per_person, aes(factor(location), revenue_per_count, fill = factor(location))
    geom_col(position = 'dodge')
```




```
revenue_by_location_per_person%>%kable()
```

location	total_revenue	count	revenue_per_count
China	86934.94	175	496.7711
Indonesia	65801.21	126	522.2318
Russia	28640.84	58	493.8076
Philippines	26957.02	55	490.1276
Brazil	18978.00	40	474.4500
France	17534.37	32	547.9491
Poland	16799.83	40	419.9957
Portugal	13215.37	28	471.9775
Ukraine	12099.07	20	604.9535
Sweden	11651.32	20	582.5660
Colombia	10220.70	21	486.7000
United States	9427.31	20	471.3655
Canada	9033.20	18	501.8444
Peru	8189.84	15	545.9893
Greece	7227.59	12	602.2992