

# HW10\_Markdown

2024-04-18

**a**

My partner is Heinrich Gonzalez. Our research question is “How do different factors affect tourism?”

**b**

Heinrich is the only other group member for this project. The grade I would give him is an A because he has been really responsive and helpful with collecting data, analysis, and overall research.

**c**

I looked from other papers on google scholar and they said that transportation, price of hotels and commodities, education, hospital facilities and languages can all affect tourism.

**d**

We found 3 datasets. One had all the general information about countries like the gdp, population, land size, etc. The second dataset had information relating to happiness so it measured social support, life expectancy, freedom, etc. The third dataset provided the tourism information about how many people come into a country in a given year. ## e

```
combined = read.csv("~/Downloads/combined.csv")
dim(combined)
```

```
## [1] 145 50
```

```
summary(combined)
```

```
##      Country      Density..P.Km2.  Abbreviation  Agricultural.Land....
## Length:145      Length:145      Length:145      Length:145
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
## Land.Area.Km2.    Armed.Forces.size    Birth.Rate    Calling.Code
## Length:145      Length:145      Min.   : 6.40  Min.   : 1.0
## Class :character Class :character 1st Qu.:10.60 1st Qu.: 66.0
## Mode  :character Mode  :character Median :17.86 Median : 250.0
##                                     Mean  :19.98 Mean  : 348.4
##                                     3rd Qu.:28.25 3rd Qu.: 421.0
##                                     Max.   :46.08 Max.   :1876.0
##
## Capital.Major.City Co2.Emissions      CPI      CPI.Change....
## Length:145      Length:145      Length:145      Length:145
## Class :character Class :character Class :character Class :character
```

```

## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
## Currency.Code Fertility.Rate Forested.Area.... Gasoline.Price
## Length:145 Min. :0.980 Length:145 Length:145
## Class :character 1st Qu.:1.630 Class :character Class :character
## Mode :character Median :2.200 Mode :character Mode :character
## Mean :2.634
## 3rd Qu.:3.400
## Max. :6.910
##
## GDP Gross.primary.education.enrollment....
## Length:145 Length:145
## Class :character Class :character
## Mode :character Mode :character
##
##
##
## Gross.tertiary.education.enrollment.... Infant.mortality Largest.city
## Length:145 Min. : 1.4 Length:145
## Class :character 1st Qu.: 4.8 Class :character
## Mode :character Median :12.8 Mode :character
## Mean :20.7
## 3rd Qu.:31.8
## Max. :84.5
##
## Life.expectancy Maternal.mortality.ratio Minimum.wage Official.language
## Min. :52.80 Min. : 2.0 Length:145 Length:145
## 1st Qu.:67.10 1st Qu.: 10.0 Class :character Class :character
## Median :74.40 Median : 39.0 Mode :character Mode :character
## Mean :72.68 Mean : 155.3
## 3rd Qu.:78.10 3rd Qu.: 186.0
## Max. :84.20 Max. :1140.0
##
## Out.of.pocket.health.expenditure Physicians.per.thousand Population
## Length:145 Min. :0.010 Length:145
## Class :character 1st Qu.:0.370 Class :character
## Mode :character Median :1.720 Mode :character
## Mean :1.916
## 3rd Qu.:3.000
## Max. :7.120
##
## Population..Labor.force.participation.... Tax.revenue....
## Length:145 Length:145
## Class :character Class :character
## Mode :character Mode :character
##
##
##
## Total.tax.rate Unemployment.rate Urban_population Latitude

```

```

## Length:145          Length:145          Length:145          Min.   :-40.90
## Class :character    Class :character    Class :character    1st Qu.:  7.54
## Mode  :character    Mode  :character    Mode  :character    Median : 23.68
##                                     Mean  : 21.84
##                                     3rd Qu.: 41.38
##                                     Max.   : 64.96
##
## Longitude          Country.Name          Country.Code          Indicator.Name
## Min.   :-106.347    Length:145           Length:145           Length:145
## 1st Qu.: -1.023     Class :character     Class :character     Class :character
## Median :  21.824    Mode  :character     Mode  :character     Mode  :character
## Mean   :  20.089
## 3rd Qu.:  46.869
## Max.   : 174.886
##
## Indicator.Code      X2016          X2017          X2018
## Length:145          Min.   :  41500   Min.   : 11000   Min.   : 14000
## Class :character     1st Qu.: 1210000   1st Qu.: 1308500   1st Qu.: 1685725
## Mode  :character     Median : 2925000   Median : 3244000   Median : 4135500
##                                     Mean  : 8744022   Mean  : 9484494   Mean  :11056899
##                                     3rd Qu.:10044000   3rd Qu.:10993100   3rd Qu.:12634000
##                                     Max.   :82682000   Max.   :86758000   Max.   :89322000
##                                     NA's   :12         NA's   :14         NA's   :27
## Overall.rank         Score          GDP.per.capita      Social.support
## Min.   :  1.00       Min.   :2.905       Min.   :0.0000      Min.   :0.000
## 1st Qu.: 40.00       1st Qu.:4.447       1st Qu.:0.6050      1st Qu.:1.048
## Median : 79.00       Median :5.358       Median :0.9400      Median :1.265
## Mean   : 78.45       Mean   :5.380       Mean   :0.8886      Mean   :1.214
## 3rd Qu.:118.00      3rd Qu.:6.167       3rd Qu.:1.1760      3rd Qu.:1.466
## Max.   :156.00      Max.   :7.632       Max.   :2.0960      Max.   :1.644
##
## Healthy.life.expectancy Freedom.to.make.life.choices Generosity
## Min.   :0.0000       Min.   :0.0000           Min.   :0.0000
## 1st Qu.:0.4240       1st Qu.:0.3560           1st Qu.:0.1080
## Median :0.6450       Median :0.4930           Median :0.1720
## Mean   :0.5956       Mean   :0.4568           Mean   :0.1803
## 3rd Qu.:0.7710       3rd Qu.:0.5800           3rd Qu.:0.2380
## Max.   :1.0080       Max.   :0.7240           Max.   :0.5980
##
## Perceptions.of.corruption
## Min.   :0.00000
## 1st Qu.:0.05075
## Median :0.08200
## Mean   :0.11199
## 3rd Qu.:0.13650
## Max.   :0.45700
## NA's   :1

```

There are 50 total variables and 145 observations. Each observation represents a different country. Some countries were omitted because the tourism data was not collected for them. 19 of the variables are doubles and the remainder are characters. However, some that are characters should be integers or doubles which we will have to change in the future.

## **f**

One method we plan on using is prediction, specifically linear regressions. We are doing this just to see if we can find high correlations among certain variables. We will also plan on doing unsupervised learning to see if there are clusters we can identify.