Project Description and Personal Preferences

Provide candidate cities and countries to relocate. The Final decision will be given after researching candidate cities and countries in detail. Social states, not low-tax, high-risk return countries. High IQ average, excellent education system. Strong institutions and other general things(democracy, media, etc.). Suitable for working remotely and entrepreneurs.

Finding Data

Required data is collected from various sources by copying and pasting to CSV files(There was only one table on each website). Web scraping will be used later.

Preparing Data with MySQL

Create a database.

```
CREATE DATABASE countryForFreelanceRemote
```

Imported CSV files.

Create the main table.

```
CREATE TABLE countries
    (city varchar(255),
    states varchar(255),
    country varchar(255),
    cost_of_living double,
    cost_of_rent double,
    health_care_index double,
    internet_mbps double,
    pollution_index double,
    safety_index double,
    taxes double
    );
```

```
ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN a varchar(255) FIRST;
```

Insert city/state/country, cost of living index and rent index to main table.

```
INSERT INTO countryforfreelanceremote.countries (a, cost_of_living, cost_of_rent)
SELECT `City`, `Cost of Living Index`, `Rent Index`
FROM countryforfreelanceremote.costs
```

Update healthcareindex of the main table on city.

```
UPDATE
    countryforfreelanceremote.countries as c
```

```
INNER JOIN
    countryforfreelanceremote.healthcare as h
ON c.a = h.city
SET
    c.health_care_index = h.`health care index`;
```

Update pollutionindex of the main table on city.

```
UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.pollution as p
ON c.a = p.City
SET
    c.pollution_index = p.`Pollution Index`;
```

Update safetyindex of the main table on city.

```
UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.safety as s
ON c.a = s.City
SET
    c.safety_index = s.`Safety Index`;
```

Split a(city/states/country) to city, states, country columns.

```
UPDATE countryforfreelanceremote.countries
SET city = (SELECT SUBSTRING_INDEX(`a`,',',1));
```

```
UPDATE countryforfreelanceremote.countries
SET country =(SELECT SUBSTRING_INDEX(`a`,',',-1));
```

Prepare states.

```
UPDATE
    countryforfreelanceremote.countries
SET states = REPLACE(`a`,`city`,"");

UPDATE
    countryforfreelanceremote.countries
SET a = REPLACE(`states`,`country`,"");

UPDATE
    countryforfreelanceremote.countries
SET states = REPLACE(`a`,',',"");
```

Trim them all.

```
UPDATE
```

```
    countryforfreelanceremote.countries
SET
    city = TRIM(`city`),
    states = TRIM(`states`),
    country = TRIM(`country`);
```

Drop a.

```
ALTER TABLE countryforfreelanceremote.countries
DROP COLUMN a;
```

Update internet_mbps of the main table on city.

```
UPDATE countryforfreelanceremote.internet
SET Country = 'Hong Kong'
WHERE Country = 'Hong Kong (SAR) '

UPDATE countryforfreelanceremote.internet
SET Country = 'Macao'
WHERE Country = 'Macau (SAR) '

UPDATE countryforfreelanceremote.internet
SET Country = TRIM(Country)

UPDATE countryforfreelanceremote.countries
SET country = 'Hong Kong'
WHERE country = 'Hong Kong (China)'

UPDATE countryforfreelanceremote.countries
SET country = 'Macao'
WHERE country = 'Macao (China)';

UPDATE countryforfreelanceremote.countries
SET country = 'Kosovo'
WHERE country = 'Kosovo (Disputed Territory)';

UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.internet as i
ON c.country = i.Country
SET
    c.internet_mbps = i.`Mbps`;
```

Update taxes of the main table on city.

```
UPDATE countryforfreelanceremote.taxes
SET Country = TRIM(Country);
```

```
ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN personal_income_taxes double;

UPDATE countryforfreelanceremote.taxes
SET `Income Tax` = REPLACE(`Income Tax`,'%','');

UPDATE countryforfreelanceremote.taxes
SET `Sales Tax` = REPLACE(`Sales Tax`,'%','');

UPDATE countryforfreelanceremote.taxes
SET `Corporate Tax` = REPLACE(`Corporate Tax`,'%','');

UPDATE countryforfreelanceremote.taxes
SET
    `Corporate Tax` = TRIM(`Corporate Tax`),
    `Sales Tax` = TRIM(`Sales Tax`),
    `Income Tax` = TRIM(`Income Tax`);

UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.taxes as t
ON c.country = t.Country
SET
    c.personal_income_taxes = t.`Income Tax`;

ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN sales_taxes double;

ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN corporate_taxes double;

UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.taxes as t
ON c.country = t.Country
SET
    c.sales_taxes = t.`Sales Tax`;

UPDATE
    countryforfreelanceremote.countries as c
INNER JOIN
    countryforfreelanceremote.taxes as t
ON c.country = t.Country
SET
```

```
    c.corporate_taxes = t.`Corporate Tax`;
```

Dropped tables.

Delete rows with missing data.

```
DELETE FROM countryforfreelanceremote.countries
WHERE
    `health_care_index` IS Null
    OR internet_mbps IS Null
    OR pollution_index IS Null
    OR safety_index IS Null
    OR personal_income_taxes IS Null;

ALTER TABLE countryforfreelanceremote.countries
DROP COLUMN taxes;
```

Finding More Data

Collected more data. Saved as CSV file.

Preparing Data with MySQL

Imported CSV files.

Create new fields on main table.

```
ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN(
    corruption double,
    economic_complexity double,
    democracy double,
    liberal_democracy double,
    innovation double,
    competitiveness double,
    labour_skills double,
    infrastructure double,
    access_to_capital double,
    openness_for_business double,
    gdp_per_capita double,
    gender_inequality double,
    gini double,
    iq double,
    press_freedom double,
    pisa double,
    public_social_exp_as_gdp double
    );
```

Update corruption of the main table on country.

```
UPDATE countryforfreelanceremote.corruption
SET Country = TRIM(Country)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.corruption as cr
ON c.country = cr.Country
SET c.corruption = cr.corruption
```

Update economic_complexty of the main table on country.

```
UPDATE `countryforfreelanceremote`.`country complexity rankings 2021`
SET `ï»¿Country` = TRIM(`ï»¿Country`)

UPDATE `countryforfreelanceremote`.`country complexity rankings 2021`
SET `ï»¿Country` = 'United States'
WHERE `ï»¿Country` = 'United States of America';

UPDATE `countryforfreelanceremote`.`country complexity rankings 2021`
SET `ï»¿Country` = 'Turkey'
where `ï»¿Country` = 'Turkiye';

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.`country complexity rankings 2021` as ec
ON c.country = ec.`ï»¿Country`
SET c.economic_complexity = ec.ECI
```

Update democracy and liberal_democracy fields of the main table on country.

```
UPDATE `countryforfreelanceremote`.`democracy`
SET `MyUnknownColumn` = TRIM(`MyUnknownColumn`)

UPDATE `countryforfreelanceremote`.`democracy`
SET `MyUnknownColumn` = 'United States'
WHERE `MyUnknownColumn` = 'United States of America';

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.`democracy` as d
ON c.country = d.`MyUnknownColumn`
SET c.democracy = d.democracy;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.`democracy` as d
ON c.country = d.`MyUnknownColumn`
SET c.liberal_democracy = d.`liberal democracy`;
```

Update entrepreneurship related fields of main table on country.

```
UPDATE countryforfreelanceremote.entrepreneurship
SET `MyUnknownColumn` = TRIM(`MyUnknownColumn`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.innovation = e.`innovation`;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.competitiveness = e.`competitiveness`;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.labour_skills = e.`labour skills`;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.infrastructure = e.`infrastructure`;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.access_to_capital = e.`access to capital`;

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.entrepreneurship as e
ON c.country = e.MyUnknownColumn
SET c.openness_for_business = e.`openness for business`;
```

Update gdp_per_capita of main table on country.

```
UPDATE countryforfreelanceremote.gdp_per_capita
SET `country` = TRIM(`country`)

UPDATE countryforfreelanceremote.gdp_per_capita
SET gdp_per_capita = REPLACE(gdp_per_capita,'$','')

UPDATE countryforfreelanceremote.gdp_per_capita
SET gdp_per_capita = TRIM(gdp_per_capita)

UPDATE countryforfreelanceremote.gdp_per_capita
SET gdp_per_capita = REPLACE(gdp_per_capita,',','')
```

```
ALTER TABLE countryforfreelanceremote.gdp_per_capita
MODIFY COLUMN gdp_per_capita double

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.gdp_per_capita as g
ON c.country = g.country
SET c.gdp_per_capita = g.`gdp_per_capita`;
```

Update entrepreneurship related fields of main table on country.

```
UPDATE countryforfreelanceremote.gender_inequality
SET `country` = TRIM(`country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.gender_inequality as g
ON c.country = g.country
SET c.gender_inequality = g.`gii`;
```

Update gini of main table on country.

```
UPDATE countryforfreelanceremote.gini
SET `country` = TRIM(`country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.gini as g
ON c.country = g.country
SET c.gini = g.`gini`;
```

Update iq of main table on country.

```
UPDATE countryforfreelanceremote.iq
SET `Country` = TRIM(`Country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.iq as g
ON c.country = g.country
SET c.iq = g.`IQ`;
```

Update media of main table on country.

```
UPDATE countryforfreelanceremote.media
SET `country` = TRIM(`country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.media as g
ON c.country = g.country
SET c.press_freedom = g.`press`;
```

Update pisa of main table on country.

```
UPDATE countryforfreelanceremote.pisa
```

```
SET `Country` = TRIM(`Country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.pisa as g
ON c.country = g.country
SET c.pisa = g.`PISA`;
```

Update social_exp_gdp of main table on country.

```
UPDATE countryforfreelanceremote.social_exp
SET `Country` = TRIM(`Country`)

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.social_exp as g
ON c.country = g.country
SET c.public_social_exp_as_gdp = g.`public_social_expenditure_as_gdp`;
```

Some corrections.

```
UPDATE countryforfreelanceremote.countries
SET pisa = '483'
WHERE country = 'Spain'

UPDATE countryforfreelanceremote.countries
SET pisa = '498'
WHERE country = 'Switzerland'

UPDATE countryforfreelanceremote.countries
SET press_freedom = '78.51'
WHERE country = 'United Kingdom'

UPDATE countryforfreelanceremote.countries
SET press_freedom = '71.22'
WHERE country = 'United States'
```

Collecting More Data

Collected innovation, competitiveness and human capital indexes. Saved as CSV files.

Preparing Data with MySQL

Imported CSV files.

Drop some columns.

```
ALTER TABLE countryforfreelanceremote.countries
```

```
DROP COLUMN innovation,
DROP COLUMN competitiveness,
DROP COLUMN labour_skills,
DROP COLUMN infrastructure,
DROP COLUMN access_to_capital,
DROP COLUMN openness_for_business;
```

Add columns.

```
ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN innovation double,
ADD COLUMN`competitiveness` double,
ADD COLUMN `human_capital` double;
```

Update competitiveness of main table on country.

```
UPDATE countryforfreelanceremote.competitiveness
SET `country` = TRIM(`country`);

UPDATE countryforfreelanceremote.competitiveness
SET `country` = 'United States'
WHERE `country` = 'USA'

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.competitiveness as g
ON c.country = g.country
SET c.competitiveness = g.`competitiveness`;
```

Update innovation of main table on country.

```
UPDATE countryforfreelanceremote.innovation
SET `country` = TRIM(`country`);

UPDATE countryforfreelanceremote.innovation
SET `country` = 'United States'
WHERE `country` = 'USA'

UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.innovation as g
ON c.country = g.country
SET c.innovation = g.`innovation`;
```

Update human_capital of main table on country.

```
UPDATE countryforfreelanceremote.human_capital
SET `country` = TRIM(`country`);

UPDATE countryforfreelanceremote.human_capital
SET `country` = SUBSTRING(`country`,2);
```

```
UPDATE countryforfreelanceremote.countries as c
INNER JOIN countryforfreelanceremote.human_capital as g
ON c.country = g.country
SET c.human_capital = g.`potential_reached`;
```

Drop tables.

Small fixes.

```
UPDATE countryforfreelanceremote.countries
SET innovation = '59.7'
WHERE country = 'United Kingdom'
```

gini*gpd_per_capita.

```
ALTER TABLE countryforfreelanceremote.countries
ADD COLUMN giniXgdp_per_capita double

UPDATE countryforfreelanceremote.countries
SET giniXgdp_per_capita = `public_social_exp_as_gdp` / `gini`
```

Exported data before deleting null rows(some countries will be deleted).

Delete null rows.

```
DELETE FROM countryforfreelanceremote.countries
WHERE
    corruption IS null OR
    economic_complexity IS null OR
    democracy IS null OR
    liberal_democracy IS null OR
    gdp_per_capita IS null OR
    gender_inequality IS null OR
    gini IS null OR
    iq IS null OR
    press_freedom IS null OR
    pISa IS null OR
    public_social_exp_as_gdp IS null OR
    innovation IS null OR
    competitiveness IS null OR
    human_capital IS null OR
    giniXgdp_per_capita IS null;
```

Web Scraping

Create scrapyproject.

```
scrapy startproject citiestolive
scrapy genspider citiespider https://www.numbeo.com/cost-of-living/
```

Added ipython shell to scrapy.cfg

Spider.

```
import scrapy
import csv

class CitiespiderSpider(scrapy.Spider):
    name = "citiespider"
    allowed_domains = ["www.numbeo.com"]
    start_urls = ["https://www.numbeo.com/cost-of-living/"]

    def parse(self, response):
        countries = response.css('div form select option::attr(value)').getall()
        for country in countries:
            country_ = country.replace(" ", "+")
            country_url = "https://www.numbeo.com/cost-of-living/country_result.jsp?country=
            yield response.follow(country_url, callback= self.parse_country_link)

    def parse_country_link(self, response):
        city_links = response.css('tbody tr td a::attr(href)').getall()
        for city_link in city_links:
            yield response.follow(city_link, callback= self.parse_city_link)

    def parse_city_link(self, response):
        quality_of_life_link = response.css('body div div div div ul li a::attr(href)').get(
        yield response.follow(quality_of_life_link, callback= self.parse_scores)

    def parse_scores(self, response):

        city_name = response.css('body div h1::text').getall()
        country_name = response.css('nav span a span[itemprop="name"]::text').getall()[1]
        cost_of_living = response.css('td[style="text-align: right"]::text').getall()[5]
        safety = response.css('td[style="text-align: right"]::text').getall()[2]
        health_care = response.css('td[style="text-align: right"]::text').getall()[3]
        pollution = response.css('td[style="text-align: right"]::text').getall()[8]

        file = open("C:/depo/repositories/best_countries_to_live/citiestolive/scraped.csv",
        writer = csv.writer(file)

        writer.writerow([city_name, country_name, cost_of_living, safety, health_care, pollu

        file.close()
```

Preparing Analysis File with Excel

Exported data from MySQL. Added to the analysis file. Text to columns. Added scraped new cities to the analysis file. Some cleaning with Excel for the scraped data.

```
Replace all "[' Quality of Life in " with ""
Replace all "']" with ""
Text to columns
=TRIM(D1)
=IF(LEN(C1)<3,C1,"")    to separate states from countries
Added headers
```

Filled missing columns of the scraped data with the existing data(Country to country match).

```
=XLOOKUP([@Country],Table1[Country],Table1[Personal Income Taxes])
=XLOOKUP([@Country],Table1[Country],Table1[Sales Taxes])
=XLOOKUP([@Country],Table1[Country],Table1[Corporate Taxes])
=XLOOKUP([@Country],Table1[Country],Table1[Corruption])
=XLOOKUP([@Country],Table1[Country],Table1[Economic Complexity])
=XLOOKUP([@Country],Table1[Country],Table1[Democracy])
=XLOOKUP([@Country],Table1[Country],Table1[Liberal Democracy])
=XLOOKUP([@Country],Table1[Country],Table1[Gender Inequality])
=XLOOKUP([@Country],Table1[Country],Table1[IQ])
=XLOOKUP([@Country],Table1[Country],Table1[Press Freedom])
=XLOOKUP([@Country],Table1[Country],Table1[Pisa])
=XLOOKUP([@Country],Table1[Country],Table1[Public Social Expenditure as GDP])
=XLOOKUP([@Country],Table1[Country],Table1[Innovation])
=XLOOKUP([@Country],Table1[Country],Table1[Human Capital])
=XLOOKUP([@Country],Table1[Country],Table1[Gini Coefficient * GDP Per Capita])
```

Copied and pasted the entire table as a value to delete formulas. Deleted cities with missing data(their countries were already deleted). Added to main sheet. Deleted dublicates.

Data Analysis With Excel

Since they were not helpful, internet speed, GDP per capita, Gini coefficient, and competitiveness data is deleted. Rearranged every criterion from 0 to 100.

```
=([@Column1]-MIN([Column1]))/(MAX([Column1])-MIN([Column1]))*100
Copy pasted as a value
Old data is deleted

=(MAX([Column1])-[@Column1])/(MAX([Column1])-MIN([Column1]))*100
Copy pasted as a value
Old data is deleted
```

```
Used these two formulas for every column.After creating a new column, cut the old column's r
```

Calculated Cons by summing taxes and costs. Calculated with summing everything - cons. Calculated pros*cons. Rearranged them from 0 to 100

Visualization with Power BI

Created cards, slicers for cities and countries, one table with every column, and bar charts for pros, cons, and pros and cons.

Results

Cities from Denmark, Finland, Sweden, Norway, Holland, Switzerland, and Germany will be analyzed in detail.