MGSC-5111-10 BUSINESS ANALYTICS FUNDAMENTALS

THE FACTORS INFLUENCING THE HAPPINESS OF PEOPLE AROUND THE WORLD

Group Members:
Anithamary Arulrayan – 0273115
Revathi Thangadurai - 0279304
Neetu Kumari - 0265402
Birajkumar Nayi – 0275418

Table of content:

1. Introduction

Various social organizations like 'Sustainable Development Solutions Network', 'The Happier way foundation', 'United Nation Organizations' etc. publish 'World Happiness Report' by collecting data from more than 100+ countries across the world. This report is gaining popularity reaching more viewership year on year since this can throw more insight into the factors affecting their happiness, living conditions of people, their priorities in life across the World and this report can help the governments and various social welfare organizations to plan for the steps to be taken for the betterment of life. (Helliwell, et al., 2022)

1.1.Motivation:

When we came across 'Happiness and Corruption 2015-2020' data in Kaggle, we the team members were curious to know

- The relation between Happiness and Corruption
- What are all the factors influencing happiness of people from different countries in the world
- On what intensity these factors are influencing the happiness
- In what way corruption can influence happiness
- Is there any way we can predict the happiness by analysis

These thoughts motivated us to consider this topic for our project and to analyze in our own way to find answers for all our above queries. (Turk, 2022)

1.2.Goals:

Our goal in this project is to analyze "Happiness and Corruption 2015-2020" data in Kaggle and to find the factors affecting global people in their happiness level, their levels of influence and to know in what way corruption affects the World Happiness Score.

1.3.Assumption:

- Conclusions and Recommendations are drawn based on analysis done for 1 year that is for 2019.
- Though it is recommended to do for many years to draw into any conclusion, the scope of this project is confined to 1 year i.e., 2019.
- One of the influencing factors "Social Support value" is 0 in the data set selected and hence it is assumed that this factor has no influence and hence omitted from the analysis.

1.4.Background Introduction:

The various factors influencing "The Happiness Score" are:

- **Happiness Score:** The score measured by surveying people from different countries on how they would rate their happiness.
- **GDP Per Capita:** Gross Domestic Product Per Capita (Countries GDP divided by its total population).
- **Family:** Quality of Family Life.
- **Health:** Life Expectancy.

- **Freedom:** Individual rights to express and make decision.
- **Generosity:** Being grateful, kind and willingness to give.
- **Government Trust:** Having confidence in Government.
- **Dystopia residual:**
  Dystopia is an imaginary country that has the world's least happy people, while residuals refer to the unexplained components that contribute to a country's happiness. We treat Dystopia Residual as a benchmark against which we may compare the evaluation of a country's happiness.t is the sum of the dystopia happiness score (1.85) i.e., score of a hypothetical country having rank lower than the lowest ranking country in the report, plus the residual value of each country which is a number left over from the normalization of the variables which cannot be explained). (Berman, 2017)
- **Social Support:** The care received from the Government and the society we live in.
- **CPI Score:** Corruption Perception Index. Higher the score lower is the corruption in the country.

2. Problem Definition:

2.1. Recognizing A Problem:

Since not all countries around the world have the same life conditions, infrastructure, government support, natural resources and economy, the individual's happiness score cannot be the same and hence there is a need to analyze what all factors can influence their happiness.

We can no way say that if a country is insufficient with facilities, they are least happy or the other way without analyzing the data.

2.2. Problem Definition:

Analyzing different factors like GDP Per Capita, Family, Health, Freedom, Generosity, Government, Dystopia Residual, Social Support and CPI Score for different countries to find out the magnitude of their influences in the people's happiness.

2.3. Challenges:

- Though it is assumed that data taken from Kaggle are reliable, we cannot be 100% sure on the reliability of the data where in the entire analysis is done.
- In the data set taken for analysis, though the data is provided from 2015 to 2020, in 2015 'social support factor' was 0 and in 2016 'Family' and 'Dystopia Residual' were given 0 and the same trend continued till 2020.
- So, it was challenging in considering the list of independent factors for analysis.
- Since it is globally accepted 'Family' plays a major role in determining individual's happiness, we, the project members decided to analyze data of 2019 because of 2 reasons namely
  - We didn't want to go with 2020 dataset though that was the latest of all, since 'Family' data is 0 for all countries and we felt that we cannot omit Family parameter.
  - Hence, we decided to analyze 2019 dataset the next latest record available wherein we have values for 'Family' and other important factors.

3. Approach and Technology

- We have approached using Descriptive and Predictive Model for solving the problem defined.
- The techniques that we have used are
    - Descriptive Statistics
    - Correlation and
    - Multiple Regression Analysis

4. Data Collection

- The dataset used for analysis is collected from one of the reliable websites www.kaggle.com "Happiness and Corruption 2015-2020"
- The dataset is available in CSV format.

4.1. Processing:

- The dataset in CSV format was uploaded in Excel.
- In 2019 dataset, 'social support' was 0 for all countries and hence we omitted that attribute in our analysis.
- The dataset with which we have done our analysis is

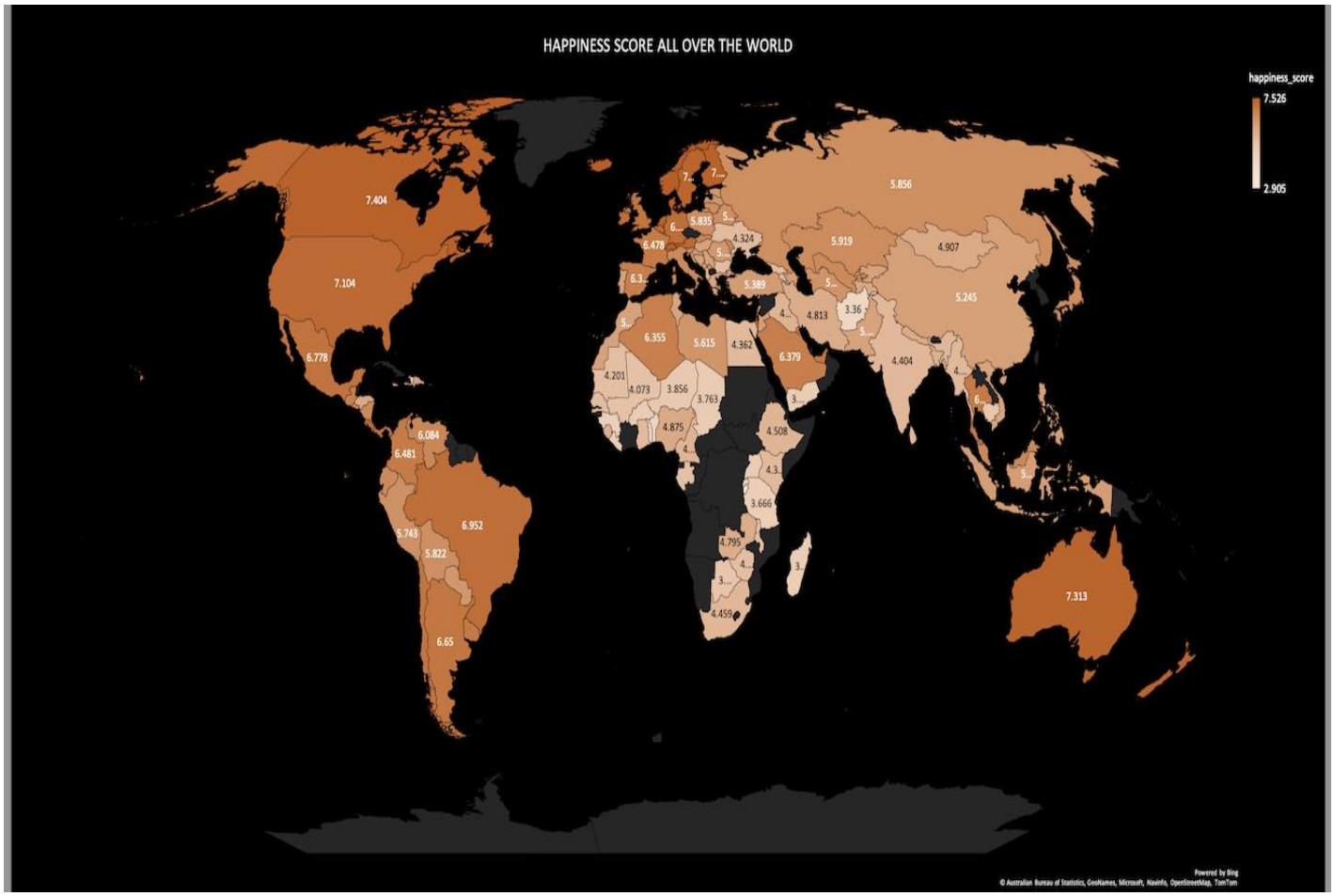
Country	happiness_score	gdp_per_capita	family	health	cpi_score	dystopia_residual	freedom	generosity	government_trust	social_support	continent	Year
Algeria	6.355	1.05266	0.83309	0.61804	35	3.40904	0.21006	0.07044	0.16157	0	Africa	2019
Mauritius	5.648	1.14372	0.75695	0.66189	52	2.20223	0.46145	0.36951	0.05203	0	Africa	2019
Libya	5.615	1.06688	0.95076	0.52304	18	2.39374	0.40672	0.17087	0.10339	0	Africa	2019
Morocco	5.151	0.84058	0.38595	0.59471	41	2.94891	0.25646	0.04053	0.08404	0	Africa	2019
Tunisia	5.045	0.97724	0.43165	0.59577	43	2.68413	0.23553	0.03936	0.0817	0	Africa	2019
Nigeria	4.875	0.75216	0.64498	0.05108	26	2.88586	0.27854	0.23219	0.0305	0	Africa	2019
Zambia	4.795	0.61202	0.6376	0.23573	34	2.58991	0.42662	0.17866	0.11479	0	Africa	2019
Sierra Leone	4.635	0.36485	0.628	0	33	3.01402	0.30685	0.23897	0.08196	0	Africa	2019
Cameroon	4.513	0.52497	0.62542	0.12698	25	2.5198	0.42736	0.2268	0.06126	0	Africa	2019
Ethiopia	4.508	0.29283	0.37932	0.34578	37	2.65614	0.36703	0.29522	0.1717	0	Africa	2019
South Africa	4.459	1.02416	0.96053	0.18611	44	1.64227	0.42483	0.13656	0.08415	0	Africa	2019
Egypt	4.362	0.95395	0.49813	0.52116	35	1.96895	0.18847	0.12706	0.10393	0	Africa	2019
Kenya	4.356	0.52267	0.7624	0.30147	28	1.88326	0.40576	0.41328	0.06686	0	Africa	2019
Ghana	4.276	0.63107	0.49353	0.29681	41	2.2002	0.40973	0.21203	0.0326	0	Africa	2019
Senegal	4.219	0.44314	0.77416	0.40457	45	1.97861	0.31056	0.19103	0.11681	0	Africa	2019
Mauritania	4.201	0.61391	0.84142	0.28639	28	1.9263	0.1268	0.22686	0.17955	0	Africa	2019
Zimbabwe	4.193	0.35041	0.71478	0.1595	24	2.4427	0.25429	0.18503	0.08582	0	Africa	2019
Malawi	4.156	0.08709	0.147	0.29364	31	2.82859	0.4143	0.30968	0.07564	0	Africa	2019
Gabon	4.121	1.15851	0.72368	0.3494	31	1.45332	0.28098	0.06244	0.09314	0	Africa	2019
Mali	4.073	0.31292	0.86333	0.16347	29	2.11087	0.27544	0.21064	0.13647	0	Africa	2019
Botswana	3.974	1.09426	0.89186	0.34752	61	0.96741	0.44089	0.12425	0.10769	0	Africa	2019
Niger	3.856	0.1327	0.6053	0.26162	32	2.09469	0.38041	0.2097	0.17176	0	Africa	2019
Chad	3.763	0.42214	0.63178	0.03824	20	2.30637	0.12807	0.18667	0.04952	0	Africa	2019
Burkina Faso	3.739	0.31995	0.63054	0.21297	40	1.87319	0.3337	0.24353	0.12533	0	Africa	2019
Uganda	3.739	0.34719	0.90981	0.19625	28	1.51416	0.43653	0.27102	0.06442	0	Africa	2019
Madagascar	3.695	0.27954	0.46115	0.37109	24	2.15075	0.13684	0.2204	0.07506	0	Africa	2019
Tanzania	3.666	0.47155	0.77623	0.357	37	1.37769	0.3176	0.31472	0.05099	0	Africa	2019
Liberia	3.622	0.10706	0.50353	0.23165	28	2.23284	0.25748	0.24063	0.04852	0	Africa	2019
Rwanda	3.515	0.32846	0.61586	0.31865	53	0.96819	0.5432	0.23552	0.50521	0	Africa	2019

Benin	3.484	0.39499	0.10419	0.21028	41	2.10812	0.39747	0.2018	0.06681	0	Africa	2019
Togo	3.303	0.28123	0	0.24811	29	2.1354	0.34678	0.17517	0.11587	0	Africa	2019
Burundi	2.905	0.06831	0.23442	0.15747	19	2.10404	0.0432	0.2029	0.09419	0	Africa	2019
Israel	7.267	1.33766	0.99537	0.84917	60	3.31029	0.36432	0.32288	0.08728	0	Asia	2019
Singapore	6.739	1.64555	0.86758	0.94719	85	1.99375	0.4877	0.32706	0.46987	0	Asia	2019
United Arab Emirates	6.573	1.57352	0.87114	0.72993	71	2.21507	0.56215	0.26591	0.35561	0	Asia	2019
Thailand	6.474	1.0893	1.04477	0.64915	36	2.5796	0.49553	0.58696	0.02833	0	Asia	2019
Saudi Arabia	6.379	1.48953	0.84829	0.59267	53	2.61482	0.37904	0.15457	0.30008	0	Asia	2019
Kuwait	6.239	1.61714	0.87758	0.63569	40	2.28085	0.43166	0.15965	0.23669	0	Asia	2019
Bahrain	6.218	1.44024	0.94397	0.65696	42	2.27405	0.47375	0.17147	0.25772	0	Asia	2019
Malaysia	6.005	1.25142	0.88025	0.62366	53	2.35384	0.39031	0.41474	0.09081	0	Asia	2019
Uzbekistan	5.987	0.73591	1.1681	0.50163	25	2.34638	0.60848	0.34326	0.28333	0	Asia	2019
Japan	5.921	1.38007	1.06054	0.91491	73	1.80584	0.46761	0.10224	0.18985	0	Asia	2019
Kazakhstan	5.919	1.22943	0.95544	0.57386	34	2.49325	0.4052	0.15011	0.11132	0	Asia	2019
Turkmenistan	5.658	1.08017	1.03817	0.44006	19	2.21489	0.37408	0.22567	0.28467	0	Asia	2019
Cyprus	5.546	1.31857	0.70697	0.8488	58	2.04497	0.29507	0.27906	0.05228	0	Asia	2019
Turkey	5.389	1.16492	0.87717	0.64718	39	2.29074	0.23889	0.04707	0.12348	0	Asia	2019
Indonesia	5.314	0.95104	0.87625	0.49374	40	2.03171	0.39237	0.56521	0.00322	0	Asia	2019
Jordan	5.303	0.99673	0.86216	0.60712	48	2.20142	0.36023	0.14262	0.13297	0	Asia	2019
Azerbaijan	5.291	1.12373	0.76042	0.54504	30	2.2735	0.35327	0.0564	0.17914	0	Asia	2019
Philippines	5.279	0.81217	0.87877	0.47036	34	2.23484	0.54854	0.21674	0.11757	0	Asia	2019
China	5.245	1.0278	0.79381	0.73561	41	2.17087	0.44012	0.04959	0.02745	0	Asia	2019
Kyrgyzstan	5.185	0.56044	0.95434	0.55449	30	2.28136	0.40212	0.38432	0.04762	0	Asia	2019
Pakistan	5.132	0.68816	0.26135	0.40306	32	3.18286	0.14622	0.31185	0.1388	0	Asia	2019
Lebanon	5.129	1.12268	0.64184	0.76171	28	2.07339	0.26228	0.23693	0.03061	0	Asia	2019
Vietnam	5.061	0.74037	0.79117	0.66157	37	1.9418	0.55954	0.25075	0.11556	0	Asia	2019
Tajikistan	4.996	0.48835	0.75602	0.53119	25	2.39106	0.43408	0.25998	0.13509	0	Asia	2019
Mongolia	4.907	0.98853	1.08983	0.55469	35	1.53586	0.35972	0.34539	0.03285	0	Asia	2019
Iran	4.813	1.11758	0.38857	0.64232	26	1.99817	0.22544	0.38538	0.0557	0	Asia	2019
Nepal	4.793	0.44626	0.69699	0.50073	34	2.32694	0.37012	0.3816	0.07008	0	Asia	2019
Bangladesh	4.643	0.54177	0.24749	0.52989	26	2.60904	0.39778	0.19132	0.12583	0	Asia	2019
Iraq	4.575	1.07474	0.59205	0.51076	20	1.81657	0.24856	0.19589	0.13636	0	Asia	2019
Sri Lanka	4.415	0.97318	0.84783	0.62007	38	0.91681	0.50817	0.46978	0.07964	0	Asia	2019
India	4.404	0.74036	0.29247	0.45091	41	2.18032	0.40285	0.25028	0.08722	0	Asia	2019
Myanmar	4.395	0.34112	0.69981	0.3988	29	1.50655	0.42692	0.81971	0.20243	0	Asia	2019
Armenia	4.36	0.86086	0.62477	0.64083	42	1.97864	0.14037	0.07793	0.03616	0	Asia	2019
Georgia	4.252	0.83792	0.19249	0.64035	56	1.87031	0.32461	0.06786	0.3188	0	Asia	2019
Cambodia	3.907	0.55604	0.5375	0.42494	20	1.31573	0.58852	0.40339	0.08092	0	Asia	2019
Yemen	3.724	0.57939	0.47493	0.31048	15	1.97295	0.2287	0.09821	0.05892	0	Asia	2019
Afghanistan	3.36	0.38227	0.11037	0.17344	16	2.14558	0.1643	0.31268	0.07112	0	Asia	2019
New Zealand	7.334	1.36066	1.17278	0.83096	87	2.47553	0.58147	0.49401	0.41904	0	Australia	2019
Australia	7.313	1.44443	1.10476	0.8512	77	2.5465	0.56837	0.47407	0.32331	0	Australia	2019
Denmark	7.526	1.44178	1.16374	0.79504	87	2.73939	0.57941	0.36171	0.44453	0	Europe	2019
Switzerland	7.509	1.52733	1.14524	0.86303	85	2.69463	0.58557	0.28083	0.41203	0	Europe	2019
Iceland	7.501	1.42666	1.18326	0.86733	78	2.83137	0.56624	0.47678	0.14975	0	Europe	2019
Norway	7.498	1.57744	1.1269	0.79579	84	2.66465	0.59609	0.37895	0.35776	0	Europe	2019
Finland	7.413	1.40598	1.13464	0.81091	86	2.82596	0.57104	0.25492	0.41004	0	Europe	2019
Netherlands	7.339	1.46468	1.02912	0.81231	82	2.70749	0.55211	0.47416	0.29927	0	Europe	2019
Sweden	7.291	1.45181	1.08764	0.83121	85	2.54734	0.58218	0.38254	0.40867	0	Europe	2019
Austria	7.119	1.45038	1.08383	0.80565	77	2.69343	0.54355	0.32865	0.21348	0	Europe	2019
Germany	6.994	1.44787	1.09774	0.81487	80	2.50931	0.53466	0.30452	0.28551	0	Europe	2019
Belgium	6.929	1.42539	1.05249	0.81959	75	2.61355	0.51354	0.2424	0.26248	0	Europe	2019
Ireland	6.907	1.48341	1.16157	0.81455	74	2.15988	0.54008	0.44963	0.29754	0	Europe	2019
Luxembourg	6.871	1.69752	1.03999	0.84542	80	2.11055	0.5487	0.27571	0.35329	0	Europe	2019
United Kingdom	6.725	1.40283	1.08672	0.80991	77	2.14999	0.50036	0.50156	0.27399	0	Europe	2019
Malta	6.488	1.30782	1.09879	0.80315	54	1.99032	0.54994	0.56237	0.17554	0	Europe	2019
France	6.478	1.39488	1.00508	0.83795	69	2.4744	0.46562	0.1216	0.17808	0	Europe	2019
Spain	6.361	1.34253	1.12945	0.87896	62	2.39663	0.37545	0.17665	0.06137	0	Europe	2019
Slovakia	6.078	1.27973	1.08268	0.70367	50	2.61065	0.23391	0.13837	0.02947	0	Europe	2019
Italy	5.977	1.35495	1.04167	0.85102	53	2.34918	0.18827	0.16684	0.02556	0	Europe	2019
Moldova	5.897	0.69177	0.83132	0.52309	32	3.38007	0.25202	0.19997	0.01903	0	Europe	2019
Russia	5.856	1.23228	1.05261	0.58991	28	2.59115	0.32682	0.02736	0.03586	0	Europe	2019
Poland	5.835	1.24585	1.04685	0.69058	58	2.20035	0.4519	0.14443	0.055	0	Europe	2019
Lithuania	5.813	1.2692	1.06411	0.64674	60	2.60525	0.18929	0.02025	0.0182	0	Europe	2019
Belarus	5.802	1.13062	1.04993	0.63104	45	2.38582	0.29091	0.13942	0.17457	0	Europe	2019
Slovenia	5.768	1.29947	1.05613	0.79151	60	1.79522	0.53164	0.25738	0.03635	0	Europe	2019
Latvia	5.56	1.21788	0.95025	0.63952	56	2.20859	0.27996	0.17445	0.0889	0	Europe	2019
Romania	5.528	1.1697	0.72803	0.67602	44	2.45184	0.36712	0.12889	0.00679	0	Europe	2019
Estonia	5.517	1.27964	1.05163	0.68098	74	1.81985	0.41511	0.08423	0.18519	0	Europe	2019
Croatia	5.488	1.18649	0.60809	0.70524	47	2.52462	0.23907	0.18434	0.04002	0	Europe	2019
Kosovo	5.401	0.90145	0.66062	0.54	36	2.80998	0.14396	0.27992	0.06547	0	Europe	2019
Serbia	5.177	1.03437	0.81329	0.6458	39	2.27539	0.15718	0.20737	0.04339	0	Europe	2019
Bosnia and Herzegovina	5.163	0.93383	0.64367	0.70766	36	2.48406	0.09511	0.29889	0	0	Europe	2019
Montenegro	5.161	1.07838	0.74173	0.63533	45	2.25531	0.15111	0.17191	0.12721	0	Europe	2019
Hungary	5.145	1.24142	0.93164	0.67608	44	1.95473	0.1977	0.099	0.04472	0	Europe	2019
Portugal	5.123	1.27607	0.94367	0.79363	62	1.53015	0.44727	0.11691	0.01521	0	Europe	2019
Greece	5.033	1.24886	0.75473	0.80029	48	2.12944	0.05822	0	0.04127	0	Europe	2019
Albania	4.655	0.9553	0.50163	0.73007	35	1.92816	0.31866	0.1684	0.05301	0	Europe	2019
Ukraine	4.324	0.87287	1.01413	0.58628	30	1.50066	0.12859	0.20363	0.01829	0	Europe	2019
Bulgaria	4.217	1.11306	0.92542	0.67806	43	1.15377	0.21219	0.12793	0.00615	0	Europe	2019
Canada	7.404	1.44015	1.0961	0.8276	77	2.70485	0.5737	0.44834	0.31329	0	North America	2019
United States	7.104	1.50796	1.04782	0.779	69	2.72782	0.48163	0.41077	0.14868	0	North America	2019
Mexico	6.778	1.11508	0.7146	0.71143	29	3.55906	0.37709	0.11735	0.18355	0	North America	2019
Costa Rica	7.087	1.06879	1.02152	0.76146	56	3.35168	0.55225	0.22553	0.10547	0	South America	2019
Brazil	6.952	1.08754	1.03938	0.61415	35	3.50733	0.40425	0.15776	0.14166	0	South America	2019

Chile	6.705	1.2167	0.90587	0.81883	67	2.95505	0.37789	0.31595	0.11451	0	South America	2019
Panama	6.701	1.18306	0.98912	0.70835	36	3.00559	0.48927	0.2418	0.08423	0	South America	2019
Argentina	6.65	1.15137	1.06612	0.69711	45	3.12985	0.42284	0.10989	0.07296	0	South America	2019
Uruguay	6.545	1.18157	1.03143	0.72183	71	2.67139	0.54388	0.18056	0.21394	0	South America	2019
Colombia	6.481	1.03032	1.02169	0.59659	37	3.17471	0.44735	0.15626	0.05399	0	South America	2019
Guatemala	6.324	0.83454	0.87119	0.54039	26	3.19863	0.50379	0.28808	0.08701	0	South America	2019
Venezuela	6.084	1.13367	1.03302	0.61904	16	2.97468	0.19847	0.0425	0.08304	0	South America	2019
El Salvador	6.068	0.8737	0.80975	0.596	34	3.22134	0.37269	0.08877	0.10613	0	South America	2019
Nicaragua	5.992	0.69384	0.89521	0.65213	22	2.82428	0.46582	0.29773	0.16292	0	South America	2019
Ecuador	5.976	0.97306	0.85974	0.68613	38	2.77366	0.4027	0.10074	0.18037	0	South America	2019
Bolivia	5.822	0.79422	0.83779	0.4697	31	2.91635	0.50961	0.21698	0.07746	0	South America	2019
Peru	5.743	0.99602	0.81255	0.62994	36	2.73117	0.37502	0.14527	0.05292	0	South America	2019
Paraguay	5.538	0.89373	1.11111	0.58295	28	2.16091	0.46235	0.25296	0.07396	0	South America	2019
Jamaica	5.51	0.89333	0.96372	0.59469	43	2.35682	0.43597	0.22245	0.04294	0	South America	2019
Dominican Republic	5.155	1.02787	0.99496	0.57669	28	1.69626	0.52259	0.21286	0.12372	0	South America	2019
Honduras	4.871	0.69429	0.75596	0.58383	26	2.29551	0.26755	0.2044	0.06906	0	South America	2019
Haiti	4.028	0.34097	0.29561	0.27494	18	2.37116	0.12072	0.47958	0.14476	0	South America	2019
Guinea	3.607	0.22415	0.3109	0.18829	29	2.15604	0.30953	0.29914	0.1192	0	South America	2019

❖ Worldwide Happiness Score Map

Scale of Happiness: Darker to lighter



4.2. Descriptive Statistics Analysis:

The below tabulation is about the Descriptive Statistical Analysis of 'Happiness Score'

Happiness Score	
Mean	**5.430704545**
Standard Error	0.100566719
Median	**5.395**
Mode	**3.739**
Standard Deviation	**1.155423638**
Sample Variance	1.335003782
Kurtosis	-0.909599872
Skewness	0.030707307
Range	**4.621**
Minimum	**2.905**
Maximum	**7.526**
Sum	716.853
Count	132

❖ **Inference from Descriptive Statistic Analysis:**

- The Average Happiness score of a country in the World is 5.43
- The Median value is 5.395 which means half of the countries in the dataset have the happiness score greater than 5.395 and half of them are lesser than 5.395
- Since Median < Mean, we can conclude that the 'Happiness Score' is positively skewed which means that there are few countries who are extremely happier.
- Since Standard Deviation is just 1.155, we can conclude that the 'Happiness Score' is clustered more around the mean.
- Mean ± 3 SD = 5.43 ± 1.155(3) results in range from 1.965 to 8.895.
    a. Since the minimum (2.905) and maximum value (7.526) of the 'Happiness Score' lies within above range, we can interpret that none of the country has significant different 'Happiness Score' from the rest.
    b. This implies that the factors listed in the data are enough to explain the 'Happiness Score'.

5. Predictive Analysis

Following are the steps taken in performing Predictive Analysis for the dataset.

5.1. Identifying Dependent Variable and Independent Variables:

- Happiness Score is the dependent variable
- GDP Per Capita, Family, Health, Freedom, Generosity, Government, Dystopia Residual and CPI Score are Independent Variables.

5.2. Correlation Analysis:

In order to reveal meaningful relationships between different metrics, using MS Excel, Correlation Analysis is done, and "Heat Map" is applied to the result for better visualization.

Correlation Analysis

	Happiness Score	GDP Per Capita	Family	Health	Freedom	Generosity	Government Trust	Dystopia Residual	CPI Score
Happiness Score	1								
GDP Per Capita	0.819096365	1							
Family	0.739623722	0.688154536	1						
Health	0.794154867	0.846743261	0.600027517	1					
Freedom	0.548935299	0.362911336	0.484460827	0.358714804	1				
Generosity	0.192139817	-0.015630943	0.126582357	0.069483068	0.404703467	1			
Government Trust	0.470321638	0.366060034	0.282452577	0.313612426	0.54126784	0.265506269	1		
Dystopia Residual	0.573716698	0.163809081	0.117497184	0.19782643	0.047743907	-0.082269965	0.055764664	1	
CPI Score	0.679867321	0.710282269	0.530854326	0.68617301	0.52389798	0.205172173	0.601615924	0.075824879	1

Inferences made from Correlation Analysis are as follows:

- 'GDP Per Capita' is having the highest correlation with 'Happiness Score' followed by 'Health', 'Family' and 'CPI Score'.
- 'Generosity' is having the least correlation with 'Happiness Score' followed by 'Government Trust' and 'Freedom'.

5.3. Multiple Regression Analysis:

- To assess the relation of all independent variables together in determining the dependent variable "Happiness Score" and their weightage, we need to perform Multiple Regression Analysis.
- From the Correlation Analysis done above, the least correlated factors 'Generosity' and 'Government Trust' are excluded from the independent variables list since their contribution is highly minimal.
- So, with the remaining factors namely 'GDP Per Capita', 'Health', 'Family', 'CPI Score' and 'Freedom' are considered as independent variables and 'Happiness Score' as the dependent variable for Multiple Regression Analysis

Multiple Regression Analysis Report

Regression Statistics	
Multiple R	0.9916
R Square	0.9833
Adjusted R Square	0.9825
Standard Error	0.1528
Observations	132

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	171.9664	28.6611	1227.3213	0.0000
Residual	125	2.9191	0.0234		
Total	131	174.8855			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.1009	0.0727	1.3886	0.1674	-0.0429	0.2447
gdp_per_capita	0.8717	0.0729	11.9507	0.0000	0.7273	1.0161
family	0.9788	0.0728	13.4445	0.0000	0.8347	1.1229
health	1.0012	0.1187	8.4344	0.0000	0.7663	1.2361
freedom	1.6551	0.1198	13.8164	0.0000	1.4180	1.8922
dystopia_residual	0.9868	0.0263	37.5646	0.0000	0.9348	1.0388
cpi_score	0.0045	0.0011	4.0730	0.0001	0.0023	0.0067

6. Project Outcome and Recommendation:

Based on the above Multiple Regression Analysis report, following **important** information can be derived.

6.1. Sample Correlation Coefficient (Multiple R Value) – The value is almost 1 which means that the independent variables (GDP Per Capita, Family, Health, Freedom, Dystopia Residual, CPI Score) and the dependent variable (Happiness Score) are on positive correlation.

6.2. The Coefficient of Determination (R^2) – The value 0.9833 indicates that 98.33% of the variation in happiness score can be explained by the values of independent variables considered for Multiple Regression.

6.3. Adjusted R-Squared – When there are a greater number of independent variables R-square value tends to show high value and hence in order to avoid overestimating the impact of more independent variables, adjusted R-squared value is calculated and evaluated.

Below tabulation explains the R-Squared and Adjusted R-squared values while incrementing the independent factors in Multiple Regression Analysis

Independent Factors	**R-Squared**	**Adjusted R-Squared**
GDP Per Capita	0.6709	0.6684
GDP Per Capita and Family	0.7297	0.7255
GDP Per Capita, Family and Health	0.7615	0.7560
GDP Per Capita, Family, Health and Freedom	0.7948	0.7883
GDP Per Capita, Family and Health, Freedom and Dystopia Residual	0.9811	0.9803
GDP Per Capita, Family and Health, Freedom, Dystopia Residual and CPI Score	0.9833	0.9825

Since **Adjusted R-squared** value has also increased and is the almost the same with minimal differences as **R-squared** value when added the independent factors, it can be concluded all these factors are in direct relation with the happiness score.

6.4. Standard Error – The average distance of observed values far from regression line is 15.28%

6.5. ANOVA (Analysis of Variance):

- Significant F value is examined to accept or reject null hypothesis.
- Null Hypothesis is defined as that "There is no relationship between dependent and independent variables" i.e., the Y value we predict from multiple regression equation is not closer to the real values.
- Since the value is almost 0, we can reject the null hypothesis. This implies that the predicted Happiness Score using the multiple regression equation is much closer to real Happiness Score.

6.6. Multiple Regression Equation – The coefficient of each independent factor signifies how much the mean of dependent factor changes given a one-unit shift in the independent variable while holding other variables in the model constant

From the analysis we can derive **point estimate** of the multiple regression equation as

HAPPINESS SCORE = 0.1009 +0.8717 (GDP PER CAPITA) + 0.9788 (FAMILY) + 1.0012 (HEALTH) + 1.6551 (FREEDOM) +0.9868 (DYSTOPIA RESIDUAL) + 0.0045 (CPI SCORE)

6.7. P-Value – Since P-value of all individual parameters are 0, we can conclude that none of their coefficient is 0.

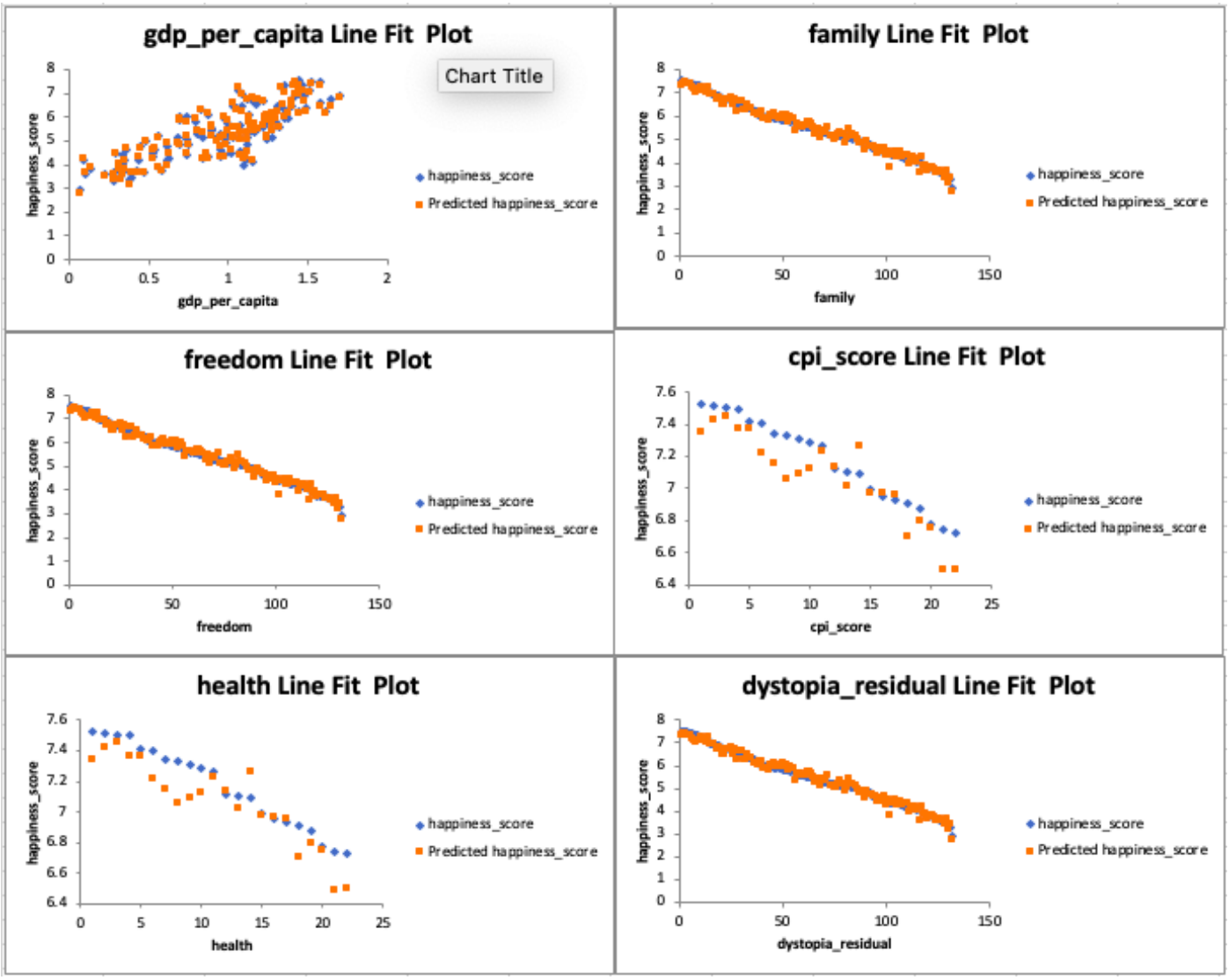
6.8. Confidence Interval:

	Coefficients	*Lower 95%*	*Upper 95%*
Intercept	0.1009	-0.0429	0.2447
gdp_per_capita	0.8717	0.7273	1.0161
family	0.9788	0.8347	1.1229
health	1.0012	0.7663	1.2361
freedom	1.6551	1.4180	1.8922
dystopia_residual	0.9868	0.9348	1.0388
cpi_score	0.0045	0.0023	0.0067

The regression equation derived from the analysis is not 100% accurate due to sampling error. The confidence interval provides a range for unknown parameters to absorb the sampling error. The above table provides the lower and upper range of coefficients of intercept and all independent factors at 95% confidence level.

6.9. Line Fit Plot:

The below line fit plot represents two sets of scatter points in blue and orange where blue points represent the actual value and the orange points represent the predicted value of the Happiness Score against each independent factor.



❖ **Interpretation of Line Fit Plots:**

- From the above Line Fit plots, we can conclude that the actual and predicted values of the Happiness Score vs GDP Per Capita, Family, Freedom, Dystopia are in best fit.
- Whereas the actual and predicted values of the Happiness Score vs CPI Score and Health are not of best fit.

7. Residuals and Residual Plots

We use Residuals and Residual Plots to determine how well a line describes our data. Residuals are the vertical distance between the regression line and the recorded data. i.e., Residual = Observed value – Predicted value

Observation	Predicted happiness_score	Residuals	Standard Residuals
1	7.346465129	0.179534871	1.202714062
2	7.428029855	0.080970145	0.542423494
3	7.45323113	0.04776887	0.320006314
4	7.369766786	0.128233214	0.859041418
5	7.369736174	0.043263826	0.289826768
6	7.222899684	0.181100316	1.213201065
7	7.152787854	0.186212146	1.247445502
8	7.063600879	0.270399121	1.811418717
9	7.093656782	0.219343218	1.469392393
10	7.123007611	0.167992389	1.125390337
11	7.230947351	0.036052649	0.241518698
12	7.136660956	-0.017660956	-0.118311726
13	7.020374529	0.083625471	0.560211674
14	7.26823459	-0.18123459	-1.214100577
15	6.974420193	0.019579807	0.131166213
16	6.968704857	-0.016704857	-0.111906765
17	6.960663946	-0.031663946	-0.212118531
18	6.704714329	0.202285671	1.355122939
19	6.795851563	0.075148437	0.503423548
20	6.75130581	0.02669419	0.178825859
21	6.489971335	0.249028665	1.668256848
22	6.494564952	0.230435048	1.543697177
23	6.71094445	-0.00594445	-0.039822202
24	6.74719752	-0.04619752	-0.309479748
25	6.836860079	-0.186860079	-1.251786042
26	6.491757112	0.081242888	0.544250613
27	6.718922777	-0.173922777	-1.165118339
28	6.237781635	0.250218365	1.676226716
29	6.636041592	-0.155041592	-1.038632233
30	6.662420384	-0.184420384	-1.235442391
31	6.250667027	0.223332973	1.496119983
32	6.269147444	0.109852556	0.735908369
33	6.522094603	-0.161094603	-1.079181689
34	6.321884635	0.033115365	0.221841667
35	6.32931515	-0.00531515	-0.035606485
36	6.151157513	0.087842487	0.588461695
37	6.155183812	0.062816188	0.420809127
38	6.055886214	0.028113786	0.188335812
39	6.168997357	-0.090997357	-0.609596342
40	6.200420609	-0.132420609	-0.887093008

104	4.478271517	-0.118271517	-0.792307453
105	4.2605342	0.0954658	0.639530692
106	4.270067133	0.053932867	0.361299271
107	4.465018152	-0.189018152	-1.266243094
108	4.29572011	-0.04372011	-0.292883442
109	4.31897264	-0.09997264	-0.669722263
110	4.339069379	-0.122069379	-0.817749544
111	3.983081804	0.217918196	1.459846089
112	4.204969778	-0.011969778	-0.08018621
113	4.231113268	-0.075113268	-0.503187951
114	4.207611159	-0.086611159	-0.580212962
115	4.051728051	0.021271949	0.142501967
116	3.583368901	0.444631099	2.978608412
117	4.234525817	-0.260525817	-1.745276905
118	3.899563008	0.007436992	0.049820825
119	3.911607171	-0.055607171	-0.372515525
120	3.703416335	0.059583665	0.399154276
121	3.790952395	-0.051952395	-0.34803198
122	3.833211432	-0.094211432	-0.631127609
123	3.774571912	-0.050571912	-0.338784045
124	3.624301356	0.070698644	0.47361414
125	3.680808959	-0.014808959	-0.099206039
126	3.674500103	-0.052500103	-0.351701104
127	3.559473446	0.047526554	0.318383022
128	3.402017928	0.112982072	0.756873171
129	3.680351584	-0.196351584	-1.315370156
130	3.176965928	0.183034072	1.226155407
131	3.406098767	-0.103098767	-0.690664364
132	2.780793727	0.124206273	0.832064715

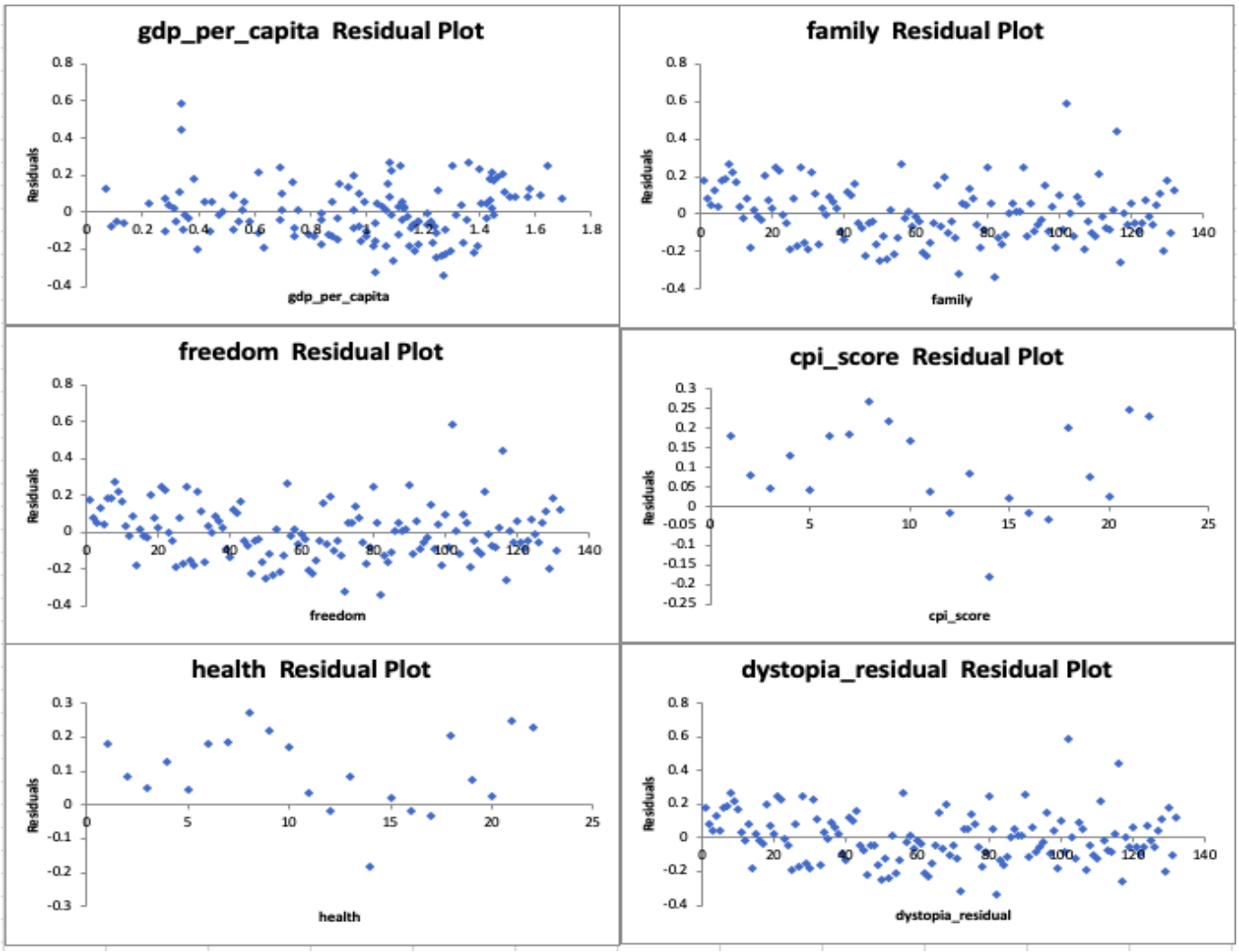
7.1. Standard Residual:

- The type of residual we often use to identify the outliers in a regression model is known as standard residuals.
- It is very similar to the kind of standardization we prefer with z scores.
- The standardized residual is the ration of the individual residual divided by the standard deviation.
- It tells about the outliers in the residuals.
- In our data range of standard residual is -2.2624678 to 3.94406257 which implies we have some outliers in our data.

7.2. Residual Plots:

Two things must be true if the regression value captures the overall pattern of the data well:

1. Residual plot should show no obvious pattern.
2. The residual should be relatively small.



7.3. Interpretation of Residual Plots:

1. All the residual plot except CPI Score and Health residual plot are randomly scattered in a way that almost half of the values are more than predicted and approximately half are less, that indicates the linear model is a good fit.
2. There are some countries which have exceptionally greater GDP Per Capita as depicted by the GDP Per Capita residual plot as some residuals are far away from the rest.
3. A larger Family or Freedom values have been recorded in few countries as displayed through the residual plot.
4. The CPI Score and Health Residual plot shows a curved pattern as most of the residues lies in the upper half which indicates that a linear model is not a good fit comparative to others.

8. Conclusion

We can conclude from the above analysis that

- Happiness score in directly influenced by GDP Per Capita, Family, Health, Dystopia Residual, Freedom and CPI Score
- The weightage is in the following order
    - Freedom
    - Health
    - Dystopia Residual
    - Family
    - GDP Per Capita
    - CPI Score
    - Government Trust
    - Generosity
- Government Trust and Generosity have negligible significance in determining the Happiness Score.
- Though Corruption Perception Index (CPI) score is also one of the significant factors in determining the Happiness of a country, the magnitude of this factor comparatively the least.
- The Happiness Score of any country can be predicted using the below equation given the following independent factors.
  HAPPINESS SCORE = 0.1009 +0.8717 (GDP PER CAPITA) + 0.9788 (FAMILY) + 1.0012 (HEALTH) + 1.6551 (FREEDOM) +0.9868 (DYSTOPIA RESIDUAL) + 0.0045 (CPI SCORE)

9. References

1. Berman, L. (2017, 12 06). *What Makes Us Happy*. Retrieved from Medium.com: https://mdium.com/@bermanlucy19/what-makes-us-happy-73659cf89a0

2. Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2022). *World Happiness Report*. Retrieved from World Happiness Report: https://worldhappiness.report/ed/2022/overview-on-our-tenth-anniversary/

3. Turk, E. (2022). *Happiness and Corruption 2015-2020*. Retrieved from Kaggle: https://www.kaggle.com/datasets/eliasturk/world-happiness-based-on-cpi-20152020

10. Appendix

Following files are attached in “World Happiness Project” folder

1. Happiness_Analysis.xlsx
2. WorldHappiness_Corruption_2015_2020.csv