

DATA ANALYTICS PRESENTATION.

Project Summary



WHAT DETERMINES HOW LONG WE LIVE?

By Favour Achuba



ME?



- A dedicated Professional with over three years' experience in healthcare recruitment, IT helpdesk and customer service.
- I'm passionate about analysing data, drawing insights, implementing solutions
- And lawn tennis!

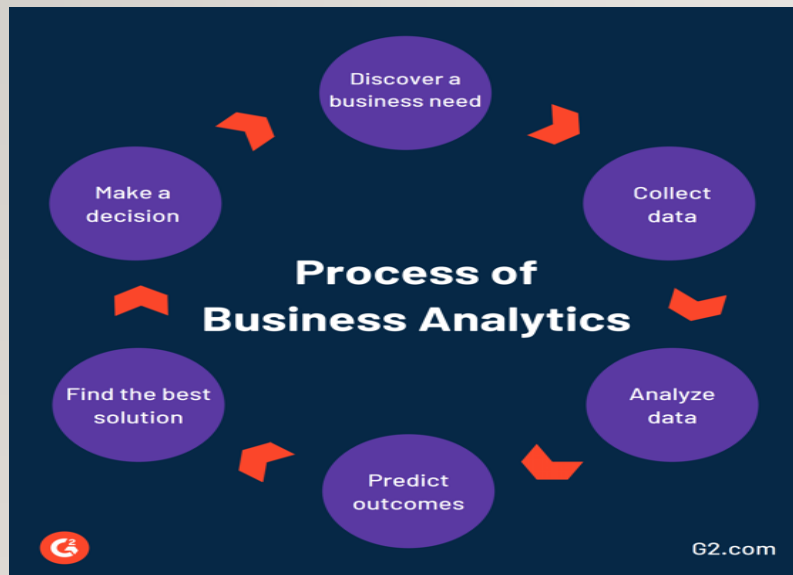
WHY DATA ANALYTICS?



- As a student, there was nothing I enjoyed more than econometrics.
- As a recruiter, I was more interested in invoicing and payroll.
- I was the HR advisor obsessed with spread sheets.
- I simply enjoy the story telling process in data!



WHAT NEXT?



I have chosen to pursue a career as a business analyst because I am more intrigued as to how data analysis impacts business processes and output...

And of course great incentives!

MY PROJECT



- What determines how long we live?
- Is it us?
- The government?
- Our Environment?
- Luck?

DATA ANALYSIS - EXCEL

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	life_expectancy			adult_mortality			infant_deaths			alcohol			percentage_expenditure	
2														
3	Mean	69.22493169		Mean	164.7964481		Mean	30.30394826		Mean	4.602860787		Mean	738.2512955
4	Standard Error	0.176006136		Standard Error	2.296983715		Standard Error	2.175632402		Standard Error	0.077360917		Standard Error	36.67514875
5	Median	72.1		Median	144		Median	3		Median	3.755		Median	64.91290605
6	Mode	73		Mode	12		Mode	0		Mode	0.01		Mode	0
7	Standard Deviation	9.523867488		Standard Deviation	124.292079		Standard Deviation	117.9265013		Standard Deviation	4.052412659		Standard Deviation	1987.914858
8	Sample Variance	90.70405193		Sample Variance	15448.5209		Sample Variance	13906.65971		Sample Variance	16.42204836		Sample Variance	3951805.483
9	Kurtosis	-0.234477394		Kurtosis	1.748860208		Kurtosis	116.0427561		Kurtosis	-0.80290922		Kurtosis	26.57338739
10	Skewness	-0.638604736		Skewness	1.174369488		Skewness	9.78696295		Skewness	0.589562528		Skewness	4.652051348
11	Range	52.7		Range	722		Range	1800		Range	17.86		Range	19479.91161
12	Minimum	36.3		Minimum	1		Minimum	0		Minimum	0.01		Minimum	0
13	Maximum	89		Maximum	723		Maximum	1800		Maximum	17.87		Maximum	19479.91161
14	Sum	202690.6		Sum	482524		Sum	89033		Sum	12630.25		Sum	2168982.306
15	Count	2928		Count	2928		Count	2938		Count	2744		Count	2938
	P	Q	R	S	T	U	V	W	X	Y	Z			
	hepatitis_B			measles			BMI			under_five_deaths				
	Mean	80.9404612		Mean	2419.59224		Mean	38.32124656		Mean	42.0357386			
	Standard Error	0.51334627		Standard Error	211.5603305		Standard Error	0.371951928		Standard Error	2.960068602			
	Median	92		Median	17		Median	43.5		Median	4			
	Mode	99		Mode	0		Mode	58.5		Mode	0			
	Standard Deviation	25.0700156		Standard Deviation	11467.27249		Standard Deviation	20.0440335		Standard Deviation	160.4455484			
	Sample Variance	628.505682		Sample Variance	131498338.3		Sample Variance	401.7632791		Sample Variance	25742.774			
	Kurtosis	2.7702594		Kurtosis	114.8599032		Kurtosis	-1.291095468		Kurtosis	109.7527951			
	Skewness	-1.9308451		Skewness	9.441331947		Skewness	-0.219311603		Skewness	9.495064657			
	Range	98		Range	212183		Range	86.3		Range	2500			
	Minimum	1		Minimum	0		Minimum	1		Minimum	0			
	Maximum	99		Maximum	212183		Maximum	87.3		Maximum	2500			
	Sum	193043		Sum	7108762		Sum	111284.9		Sum	123501			
	Count	2385		Count	2938		Count	2904		Count	2938			

EXCEL – CORRELATION RESULTS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
41	We have now had an overview of the data and spotted some interesting patterns in the data but we do not have a firm understanding of how variables in this dataset affect each other or more importantly how they impact life expectancy .														
42	Because we are focused on how the other variables correlate with life expectancy, we will focus on the first column only.														
43	life_expectancy	adult_mortality	infant_deaths	alcohol	centage_expendit	hepatitis_B	measles	BMI	under_five_deaths	polio	tal_expenditui	diphtheria	HIV_AIDS	GDP	
44	life_expectancy	1													
45	adult_mortality	-0.696359314	1												
46	infant_deaths	-0.196557177	0.078756012	1											
47	alcohol	0.404876761	-0.195848196	-0.115637677	1										
48	percentage_expenditure	0.381863503	-0.242859528	-0.085612222	0.341285313	1									
49	hepatitis_B	0.256761948	-0.162476325	-0.223566281	0.087548711	0.016273693	1								
50	measles	-0.157585804	0.031176412	0.501128342	-0.051826674	-0.056595677	-0.120529372	1							
51	BMI	0.567693548	-0.387016784	-0.227278888	0.33040846	0.228699753	0.150379532	-0.17597706	1						
52	under_five_deaths	-0.222529116	0.094146127	0.996628882	-0.112370397	-0.087852306	-0.233126251	0.507808707	-0.23766852	1					
53	polio	0.465555806	-0.274822815	-0.170688559	0.221733797	0.147259463	0.486170773	-0.13616601	0.284568764	-0.188720213	1				
54	total_expenditure	0.218086374	-0.115280689	-0.128616342	0.29694156	0.174419689	0.058280304	-0.10624059	0.242502604	-0.130148312	0.137330249	1			
55	diphtheria	0.479494864	-0.275131358	-0.175171496	0.222020171	0.143624426	0.611494949	-0.14188194	0.283147336	-0.195668288	0.673553321	0.152753524	1		
56	HIV_AIDS	-0.556556253	0.523820508	0.025231318	-0.048844563	-0.097856819	-0.112675448	0.030898718	-0.24371653	0.038061512	-0.15955954	-0.00138884	-0.164860095	1	
57	GDP	0.461455193	-0.296049318	-0.108427363	0.354712086	0.899372641	0.083903212	-0.07646605	0.301557394	-0.112081253	0.211975566	0.138364222	0.200665557	-0.13649082	
58	population	-0.021538108	-0.013646972	0.556801332	-0.035252342	-0.025661888	-0.123320952	0.265966087	-0.07230102	0.544422649	-0.03854025	-0.07966184	-0.028443781	-0.02785429	
59	The results indicate the following:														
60	adult_mortality mortality has a strong negative correlation with life expectancy. This means if adult mortality increases, life expectancy is expected to fall.														
61	Infant mortality,measles, Aids and population also have a negative correlation with life expectancy. A rise in any of the afore mentioned is expected to affect life expectancy negatively.														
62	percentage_expenditure on health, polio immunization, diphtheria immunization coverage and gdp per capita all have a psitive correlation with life expectancy indicating that as these variables increase, life expectancy should increase as well.														

DATA ANALYSIS - SQL

Query 1 Data-bootcamp-project-3

Limit to 1000 rows

```
38
39 • Select Avg(life_expectancy) as "Avg_LE", Avg(total_expenditure) as "Avg_tot_exp", Avg(income_composition_of_resources) as "Avg_inc_comp",
40 Avg(alcohol) as "Avg_alcohol", aVG(Schooling) as "avg_sch", Avg(BMI) as "Avg_BMI"
41 from project.data
42 where year BETWEEN 2000 AND 2007
43 AND status = "developing";
44
45 • Select Avg(life_expectancy) as "Avg_LE", Avg(total_expenditure) as "Avg_tot_exp", Avg(income_composition_of_resources) as "Avg_inc_comp",
46 Avg(alcohol) as "Avg_alcohol", aVG(Schooling) as "avg_sch", Avg(BMI) as "Avg_BMI"
47 from project.data
48 where year BETWEEN 2008 AND 2015
49 AND status = "developing";
50
```

Result Grid

	Avg_LE	Avg_tot_exp	Avg_inc_comp	Avg_alcohol	avg_sch	Avg_BMI
▶	65.58981788079468	5.279594370860922	0.5081382450331132	3.5829470198675586	10.076655629139076	32.96307947019871

Result Grid

	Avg_LE	Avg_tot_exp	Avg_inc_comp	Avg_alcohol	avg_sch	Avg_BMI
▶	68.06962233169128	5.008768472906394	0.5989802955665029	2.911256157635484	11.293185550082091	36.89811165845651

TREND ANALYSIS – SQL (DEVELOPED)

Query 1 Data-bootcamp-project-3 x

Limit to 1000 rows

```
56 • Select Avg(life_expectancy) as "Avg_LE", Avg(total_expenditure) as "Avg_tot_exp", Avg(income_composition_of_resources) as "Avg_inc_comp",
57 Avg(alcohol) as "Avg_alcohol", aVG(Schooling) as "avg_sch", Avg(BMI) as "Avg_BMI"
58 from project.data
59 where year BETWEEN 2000 AND 2007
60 AND status = "developed";
61
62 • Select Avg(life_expectancy) as "Avg_LE", Avg(total_expenditure) as "Avg_tot_exp", Avg(income_composition_of_resources) as "Avg_inc_comp",
63 Avg(alcohol) as "Avg_alcohol", aVG(Schooling) as "avg_sch", Avg(BMI) as "Avg_BMI"
64 from project.data
65 where year BETWEEN 2008 AND 2015
66 AND status = "developed";
67
68 -- Just as we see in the case of developing nations, we can see an increase in average life expectancy in more recent years in developed counterparts
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: IA

	Avg_LE	Avg_tot_exp	Avg_inc_comp	Avg_alcohol	avg_sch	Avg_BMI
▶	78.10312499999999	7.231445312500001	0.7564687499999997	10.096523437500005	14.008203124999994	50.062890624999994

Result Grid | Filter Rows: | Export: | Wrap Cell Content: IA

	Avg_LE	Avg_tot_exp	Avg_inc_comp	Avg_alcohol	avg_sch	Avg_BMI
▶	80.29257812499999	6.932382812500003	0.7886679687500002	8.482148437500003	14.711718750000006	53.5449218749999986

PREVALENCE OF LIFE THREATENING DISEASE

Query 1 Data-bootcamp-project-3

Limit to 1000 rows

```
-- 2. What is the prevalence of factors and life threatening disease that theoretically tend to impact life expectancy negatively.
-- what is the prevalence in both developed and developing groups?

74
75
76
77 • Select Sum(adult_mortality) as "sum_adul_mort", sum(infant_deaths) as "sum_infant_mort", Sum(measles) as "sum_measles",
78    sum(hiv_aids) as "sum_aids", Sum(thinness_1_to_19_years) as "sum_adol_thinness"
79    from project.data
80    where status = "developing";
81
82 • Select Sum(adult_mortality) as "sum_adul_mort", sum(infant_deaths) as "sum_infant_mort", Sum(measles) as "sum_measles",
83    sum(hiv_aids) as "sum_aids", Sum(thinness_1_to_19_years) as "sum_adol_thinness"
84    from project.data
85    where status = "developed";
86
```

Result Grid

	sum_adul_mort	sum_infant_mort	sum_measles	sum_aids	sum_adol_thinness
▶	441725	88268	6853271	5067.1000000000017	13378.300000000003

Result Grid

	sum_adul_mort	sum_infant_mort	sum_measles	sum_aids	sum_adol_thinness
▶	40799	765	255491	51.20000000000046	676.2000000000006

POPULATION AND LE

Query 1 Data-bootcamp-project-3

Limit to 1000 rows

```
114 • Select Avg(population) as "sample_avg" from project.data; -- ('9923150.2805')
115
116 • Select status, Avg(life_expectancy) as "avg_Le"
117   From project.data
118   Where population <= 9923150.2805
119   group by status;
120
121 • Select status, Avg(life_expectancy) as "avg_Le"
122   From project.data
123   Where population > 9923150.2805
124   group by status;
125
126 -- the results indicate that population density affects developed/developing countries differently.
```

Result Grid

	status	avg_Le
▶	Developing	67.12223873204564
	Developed	79.11930585683294

Result Grid

	status	avg_Le
▶	Developing	65.40909090909095
	Developed	79.9078431372549

GDP PER CAPITA AND LE



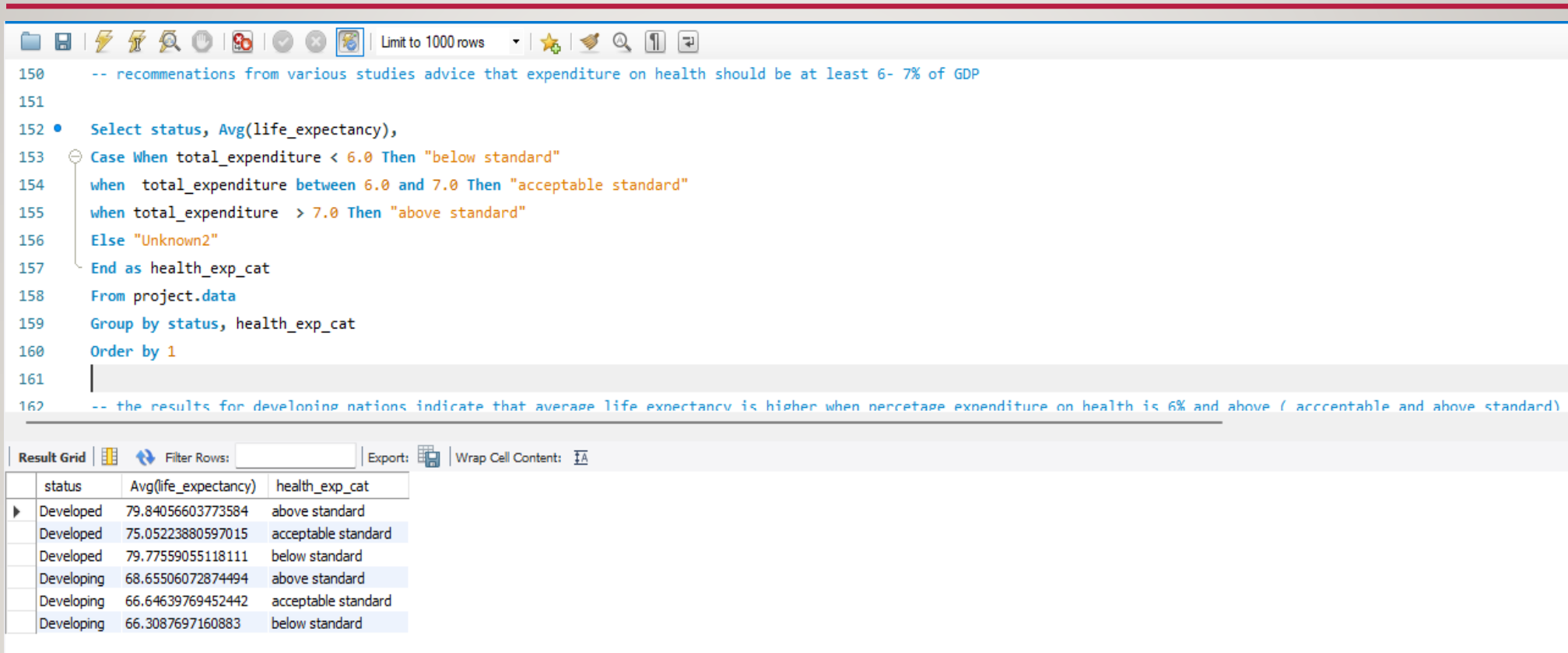
The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
135
136 • Select status, Avg(life_expectancy),
137     Case When GDP <= 1025 Then "low_income"
138     When GDP between 1026 and 12475 Then "Middle_income"
139     When GDP >=12476 Then "High_income"
140     ELSE "Unknown"
141     End as "Gdp_cat"
142 From project.data
143 Group by status, gdp_cat
144 Order by 1;
145
146 -- based on the results, it is clear that both in developed and developing worlds, countries with a high GDP per capita have a higher life expectancy
147 -- whereas. countries with low income per capita have a lower average life expectancy compared to middle and high per capita counterparts.
```

Below the query editor, there is a "Result Grid" section with a table showing the results of the query. The table has four columns: status, Avg(life_expectancy), and Gdp_cat. The results are as follows:

	status	Avg(life_expectancy)	Gdp_cat
▶	Developed	80.40042735042735	High_income
	Developed	77.58947368421053	low_income
	Developed	78.49508196721311	Middle_income
	Developing	76.64853801169592	High_income
	Developing	63.38770491803278	low_income
	Developing	70.06363636363635	Middle_income

EXPENDITURE ON HEALTH AND GDP



The screenshot displays a data analysis interface with a SQL query editor and a result grid. The query is as follows:

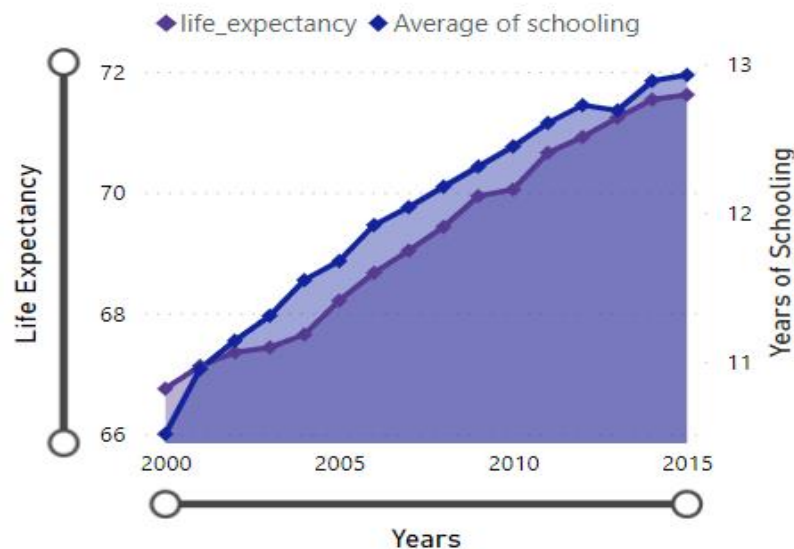
```
150 -- recommendations from various studies advice that expenditure on health should be at least 6- 7% of GDP
151
152 • Select status, Avg(life_expectancy),
153   Case When total_expenditure < 6.0 Then "below standard"
154   when total_expenditure between 6.0 and 7.0 Then "acceptable standard"
155   when total_expenditure > 7.0 Then "above standard"
156   Else "Unknown2"
157 End as health_exp_cat
158 From project.data
159 Group by status, health_exp_cat
160 Order by 1
161
162 -- the results for developing nations indicate that average life expectancy is higher when percentage expenditure on health is 6% and above ( acceptable and above standard)
```

Below the query editor, the 'Result Grid' shows the following data:

	status	Avg(life_expectancy)	health_exp_cat
▶	Developed	79.84056603773584	above standard
	Developed	75.05223880597015	acceptable standard
	Developed	79.77559055118111	below standard
	Developing	68.65506072874494	above standard
	Developing	66.64639769452442	acceptable standard
	Developing	66.3087697160883	below standard

Life Expectancy Trends and Triggers

Life expectancy and schooling Trend



Share of GDP per Capita by status

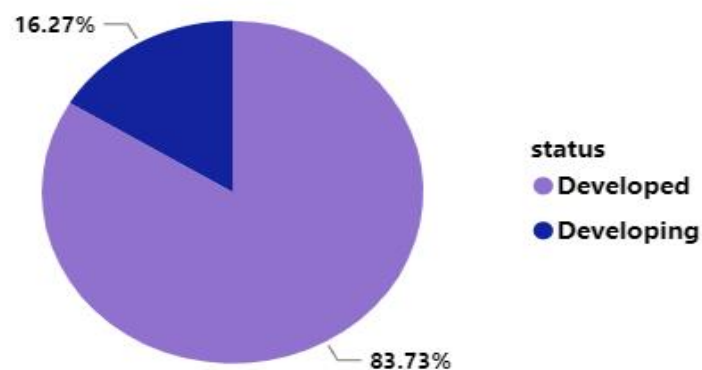


Figure 1 is a line graph showing the trend of life expectancy and years of schooling from 2000-2015. This shows that both life expectancy (7.29% increase) and schooling (22.97% increase) follow an upward trend as the years progress.

Figure 2 is a scatterplot showing the relationship between life expectancy and alcohol intake. The result indicates a negative relationship exists. Such that if there is an increase in alcohol intake, a fall in life expectancy is expected.

Figure 3 shows the share of GDP by status. here we find the developed nations account for 83.73% of the total share of GDP in the sample.

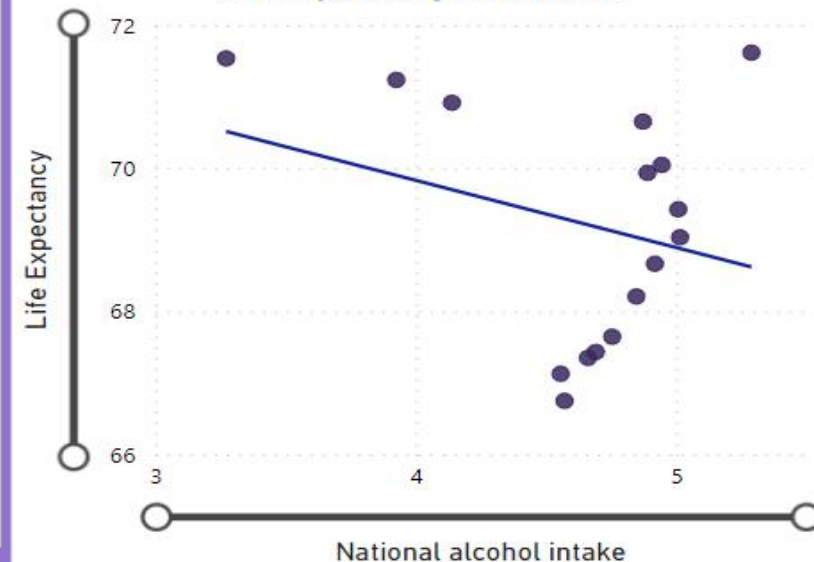
Figure 4 is an area map showing concentration of measles and aids. results shows a high prevalence in Africa, with Swaziland accounting for 10.30% of Aids in the region.

Figure 5 is a key influence visual between life expectancy and population. this indicates that when population is less than 62181 million, life expectancy increases by 3.36 years. All the results discussed support our findings from SQL analysis.

Prevalence of Measles and Aids



Life Expectancy VS Alcohol



LE and population

Key influencers

What influences life_expectancy to

Increase

When...

...the average of life_expectancy increases by

population is 62181 or less

3.36



- The trend of life expectancy, schooling, income composition, BMI has increased overtime, while alcohol intake has fallen in both groups.
- Increase in population affects life expectancy negatively in developing countries and has a positive correlation in developed countries.
- Higher GDP per capita correlates with a higher life expectancy in both groups.
- Expenditure on health that is deemed above or at required standard correlates with a higher average life expectancy in developing countries. But in developed nations, avg LE only improves when expenditure on health is above 7% of GDP

TAKE AWAY EDUCATIONALLY



PERSONALLY?

Independent Learning



CAREER WISE

**DON'T JUST
DO WHAT
YOU ARE TOLD,
DO WHAT
NEEDS
TO BE DONE.**

Face your problems head-on.
Do what you have to do to
take care of it. Develop a
good work ethic.

George Chuvalo

PeoplesQuotes.com

THANK YOU