

# Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital

Krystian Matusz

## ABSTRACT

Healthcare industry is always very demanding and opening a new Hospital is absolutely always a big challenge. There is a lot of different people, illnesses and things that you as the decision maker, planner, enabler or even GP have to face on. There is also multitude scenarios that you need to consider. Each of them can be critical for making the final, rational decision about the right location of the new hospital that will save peoples life. You need to precisely know how to find the best place and you know that the margin of error in current demanding and difficult market is absolutely minimal. It's about people's life.

This paper will present usage of public data to support making the critical business decision on the new Hospital location. The described framework is intended to help make smart(er) decisions about planning and network development. It is also intended to help better understand the demographics and local population data. The implementation will use the powerful SAS® procedures like PROC SQL, FREQ, CONTENTS, PRINT, FORMAT, SAS 4GL, SAS Macro Programming and SAS ODS. An Irish CENSUS data will be used in order to show the site selection by combining demographic data, people life stage, crime ratio, local trends, socio-demographic profiles, health care insights and behavioral - lifestyle patterns. The presented conjunction of the patient, demographic and "consumer" expenditure data allows to build the view of the potential needs, to determine an opportunities and potential market-fit.

## INTRODUCTION

Opening new Hospital is a big challenge for many organizations. There are multitude risk factors and unknown level of uncertainty. In such demanding and competitive market conditions you need to be very lucky or you should have detailed, precise information about market conditions and its performance. It's very important for you as the business owner to have all of this information. In this paper I want to present the second, data-driven approach which mostly based on the publicly information that is available for you all the time. My approach presented here is universal and applicable everywhere. The subject of this paper it to empower people (C-suite, execs, business decision makers, advisors, analysts, data scientists) in facing this tricky type of business challenge.

As a safe harbor notice I should mentioned that there are no golden rules and crystal ball so do not treat my approach as the golden rule that always works. It is rather (and that's my intention) a universal applicable framework that you can adjust to your organization purpose and needs. The final success, CX and ROI often depends on the data that you have, the market that you operate and there are multitude different factors involved here. There is also the second goal of my research presented which is to show the possibilities and raise your awareness that you can always refers to Advanced Analytics and hard data that really can help in the trickiest situations. It can save your time, your money and in this case, you can also save people's life. In my research I am going to deep dive into demographic data about population, their life stage, affluence, economic status, industry. As you know, for data scientist like me every detail really matters. That's why I dig into every piece of information that I found and that can be valuable in building econometric model. It's all about knowing and understanding population and demographics better. You may be familiar with the KYC/KYP (Know Your Customer/Know Your Patient) paradigm but you should also be familiar with characteristic of the population as a whole.

## CRISP-DM AS THE BEST METHODOLOGY OF RESEARCH

Due to challenging topic that, I used CRISP-DM<sup>1</sup> as the best **C**Ross-Industry Process for **D**ata **M**ining project. This methodology provides a well-structured approach to planning and executing the successful advanced analytics research and data mining project like this one. It is also the most commonly used approach that data mining experts use to tackle the business problems.

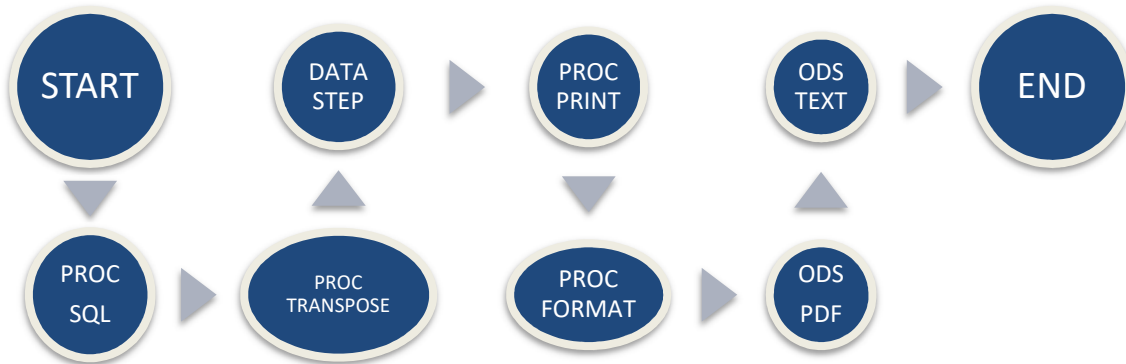
---

<sup>1</sup> <http://crisp-dm.eu/reference-model/>

## THE PROGRAM (WORK)FLOW AND ARCHITECTURE FRAMEWORK

The project is complicated in terms of data sources, data transformation, data integration, mapping and Advanced Analytical part. That's why it was necessary to build and use an appropriate structure approach for the whole implementation. Good programming practice is to distinguish the view (report, table, presentation) logic from the backend (data) layer. I made the decision to map programming design pattern (MVC) from Java EE into my project from source code (SAS 4GL and Macro Programming) perspective. This will made my solution more universal, especially that the goal was to maintain high level of automation (auto-validation, macro programming, unit testing, assertions) etc. Thanks this, changes made in backend (on the data sources level – for instance when the new CENSUS data will be published in the future) won't have impact on the final report shape using on frontend. You can recognize DAO (Data Access Object) layer as an independent element in the backend which is separated from the frontend. Because the picture is worth a thousand words and meetings, below is the data visualization diagram that presents complete structure and flow of implemented back-end and front-end logic in a high level of my project:

**Figure 1 is a schema of Program Flow**



**Figure 1. Schema of Program Flow**

As depicted above, the final schema presents the source code design pattern. Most of back-end, data preparation steps consists of querying the data (PROC SQL), checking and validation the data (MEANS, FREQ, CONTENTS), , transformation using PROC TRANSPOSE, imputation (PROC MI, PROC MIANALYZE) and mapping etc. The front-end part presents findings in an appropriate way based on analyzed subjects mentioned before. The output is not only set of statistics but also a structure way to show the key insights that came from analyzed hard data. Data visualization has been prepared thanks PROC FORMAT and PROC PRINT routines. Presented research project consists of fourteen key steps implemented from ground up in SAS Enterprise Guide and 4GL. Each step consists of approximately 1000 lines of source code (with comments included) written manually.

## THE STRUCTURE OF GEOGRAPHY OF IRELAND

Politically, Ireland is divided between the Republic of Ireland (officially named Ireland), which covers five-sixths of the island, and Northern Ireland, which is part of the United Kingdom, in the northeast of the island. Geographically, Ireland is the second-largest island of the British Isles, the third-largest in Europe, and the twentieth-largest on Earth.<sup>2</sup> The new edition of Census statistics, determined the following geography levels and their statistical frequency:

GEOGRAPHY LEVEL	STATISTICAL FREQUENCY
PROVINCES	4
GAELTACHTS	7
NUTS3 REGIONS	8
LIMISTÉIR PLEANÁLA TEANGA	26
COUNTIES	31
CONSTITUENCIES	40
MUNICIPAL DISTRICTS	95
LOCAL ELECTORAL AREAS	137

<sup>2</sup> <http://en.wikipedia.org/wiki/Ireland>



Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital, continued

	name	address	eircode	address_v2	nr_words	last_word
1	Lusk	Main Street, Lusk, Co. Dublin.		main street, lusk...	5	dublin
2	Dowra	Drumnafinnila, Dowra, Co. Leitrim		drumnafinnila, do...	4	leitrim
3	Clonakilty	Primary Care Centre Hospital Grounds, Scartagh, Clonakilty, Co. Cor...		primary care cen...	9	cork
4	Hartstown	Cherryfield Lawns, Hartstown, Dublin 15		cherryfield lawns...	4	dublin
5	The Ballagh	Ballymurry, The Ballagh, Enniscorthy, Co. Wexford.		ballymurry, the b...	6	wexford
6	Milltown	Bridge Street, Milltown, Co. Kerry		bridge street, mill...	5	kerry
7	Butler Centre	Butler Community Centre, Saint John's Park, Waterford.		butler community...	7	waterford
8	Carigeen	Church Hall, Carigeen, Co. Kilkenny.		church hall, carig...	5	kilkenny
9	Ferrybank	Community Hall, Ferrybank, Waterford.		community hall, f...	4	waterford
10	Kill	Community Hall, Kill, Co. Waterford.		community hall, ...	5	waterford
11	Passage East	Community Hall, Passage East, Co. Waterford.		community hall, ...	6	waterford
12	Emly	Galtee View, Emly, Co. Tipperary.		galtee view, emly...	5	tipperary
13	Kilsheelan	Chapel Road, Kilsheelan, Clonmel, Co. Tipperary.		chapel road, kils...	6	tipperary
14	Ballymahon	Ballymahon, Co. Longford.		ballymahon, co l...	3	longford
15	Muirhevnamore	Muirhevnamore, Co. Louth.		muirhevnamore, ...	3	louth
16	Oylegate	Oylegate, Enniscorthy, Co. Wexford.		oylegate, ennisc...	4	wexford

**Table 2. The structure of the file with list of the hospitals**

You can clean, standardize and extract that using DataFlux or any other tool to do that, but I have done that manually. That allows me to have 100% responsibility and control on my code, its behavior and be sure that I do not lose any pieces of information (entropy level). Below the partial logic and my source code snippet:

```
data EXTRNL.HEALTH_CENTRES_IRELAND; set EXTRNL.HEALTH_CENTRES_IRELAND;
  address_v2 = lowercase(compress(address, '1234567890.+~'));
  nr_words = countw(address_v2);
  last word = scan(address_v2, nr_words); run;
```

As you can see in the column address\_v2, nr\_words and last\_word have been derived but thanks manual sanitization (compress) and standard functions I successfully extracted the proper name of the county to be integrated in appropriate level. Also very useful was group processing (first and last attributes) but when you work with really big data sets, remember that it has to be sorted properly before. During the preparation phase and acquisition, cleaning, standardization and integration activities, I decided to write wrapper with a set of assertions that automatically checks the basic geography level stats (level by level). Below an example of macro structure and its invocation for a county level:

```
OPTIONS NOMLOGIC NOMPRINT NOSYMBOLGEN;
%macro check_how_many(level, title_lbl);
title "&title_lbl => Assertion: validation of count(distinct(&level.))";
proc sql threads stimer;
  select count(distinct(&level.))
  from CENSUS.CENSUS_SA_GEO_AND_KPIS ;
quit;title;
%mend check_how_many;
%check_how_many(CENCSO_SA_GEO_COUNTY_ID , ' [1] COUNTIES_LEV {31}');
%check_how_many(CENCSO_SA_GEO_COUNTY , ' [1] COUNTIES_LEV {31}');
OPTIONS MLOGIC MPRINT SYMBOLGEN;
```

I intentionally turn off the MLOGIC MPRINT SYMBOLGEN options to have a clean log file to analyze. Additionally, several sources consists of the long list of columns on a low level (Small Area or Electoral Districts level). Therefore in order to deal with efficient way of building the aggregates, I decided to implement another solution keeping in mind the level of automation that wanted to achieve:

```
PROC CONTENTS data=EXTRNL.ESB_CONNECTIONS_BY_AREA_MONTHLY
  out=work.ESB_CONNECTIONS_MONTHLY_FLDS (keep=name); run;
DATA work.ESB_CONNECTIONS_BY_AREA_MONTHLY;
  SET work.ESB_CONNECTIONS_MONTHLY_FLDS;
  SQL LABELS = CATT('sum(ESB_CONN.', name, ') as ESB_CONN ', name, ', '); run;
```

I am listing all the column names from the target data and then produce the second column that represents part of SQL query built dynamically - on fly as the code is running (no hard coded names, labels etc.). After building the

Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital, continued

whole ABT table, thanks Advanced Analytical approach (Spearman/Pearson correlation coefficient - PROC CORR), the final subset of parameters has been selected to the further analysis. Additionally the experiments were conducted with Kendall Tau-b Correlation Coefficient. Based on that subset of metrics, sixteen tables with the key actionable insights have been derived as the output of this research and analysis.

## COUNTRY MACROECONOMIC VIEW AND INSIGHTS

Name of the County	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
<b>Cork City</b>	125 657	21%	64%	84%	.	616	1 770	800	509	1 174
<b>Clare</b>	118 817	20%	90%	87%	.	2 970	8 487	6 989	2 640	7 921
<b>Cork County</b>	417 211	18%	86%	90%	.	2 045	5 876	2 657	1 689	3 899
<b>Cavan</b>	76 176	19%	91%	87%	76 176	2 240	6 925	7 618	3 047	2 930
<b>Carlow</b>	56 932	18%	83%	86%	.	2 588	7 117	4 067	2 711	8 133
<b>Dublin City</b>	554 554	17%	57%	83%	50 414	1 206	5 088	1 880	869	8 944
<b>Donegal</b>	159 192	21%	96%	86%	159 192	2 745	7 236	6 921	3 184	3 702
<b>Dún Laoghaire-Rathdown</b>	218 018	21%	79%	90%	.	.	.	.	.	.
<b>Fingal</b>	296 020	13%	74%	88%	.	.	.	.	.	.
<b>Galway City</b>	78 668	16%	58%	86%	.	803	1 788	1 405	672	2 070
<b>Galway County</b>	179 390	20%	93%	88%	.	1 831	4 077	3 203	1 533	4 721
<b>Kildare</b>	222 504	14%	78%	90%	222 504	3 272	9 674	5 563	2 747	13 088
<b>Kilkenny</b>	99 232	20%	89%	88%	99 232	2 919	5 513	6 202	3 969	2 919
<b>Kerry</b>	147 707	23%	91%	86%	147 707	2 462	6 154	4 220	2 345	3 517
<b>Longford</b>	40 873	20%	93%	85%	.	2 271	10 218	6 812	2 044	3 406
<b>Louth</b>	128 884	17%	84%	86%	128 884	2 302	9 914	5 858	2 864	9 914
<b>Limerick City and County</b>	194 899	19%	82%	86%	97 450	2 293	7 496	4 872	1 949	5 268
<b>Leitrim</b>	32 044	23%	99%	86%	.	2 289	5 341	8 011	3 204	2 465
<b>Laois</b>	84 697	16%	85%	87%	84 697	3 025	14 116	8 470	2 921	4 982
<b>Meath</b>	195 044	15%	84%	90%	195 044	3 751	10 265	6 501	4 063	13 932
<b>Monaghan</b>	61 386	19%	90%	88%	61 386	3 069	8 769	5 581	3 837	5 581
<b>Mayo</b>	130 507	24%	98%	86%	130 507	2 417	5 932	4 500	1 764	4 350
<b>Offaly</b>	77 961	19%	90%	86%	77 961	2 887	7 087	7 087	2 784	5 569
<b>Roscommon</b>	64 544	23%	97%	87%	64 544	3 586	4 034	8 068	2 391	4 034
<b>South Dublin</b>	278 767	16%	79%	87%	.	.	.	.	.	.
<b>Sligo</b>	65 535	22%	91%	86%	65 535	2 521	9 362	5 461	1 771	5 041
<b>Tipperary</b>	159 553	21%	91%	87%	79 777	2 346	5 147	5 147	2 751	4 199
<b>Waterford City and County</b>	116 176	20%	88%	87%	116 176	2 640	9 681	4 006	2 702	5 532
<b>Westmeath</b>	88 770	18%	84%	87%	88 770	2 114	5 918	5 548	2 336	5 918
<b>Wicklow</b>	142 425	18%	85%	89%	.	3 096	6 782	4 748	3 237	8 378
<b>Wexford</b>	149 722	20%	91%	87%	149 722	3 186	8 318	7 130	3 565	8 318

**Table 3. Country Macroeconomic View**

Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital, continued

Column descriptors:

- [1 ] **Population** number represents people living in the area
- [2 ] Demographic **Elderly Factor**
- [3 ] Demographic **Load Factor**
- [4 ] Health Performance Index – percentage of people with **good health**
- [5 ] Number of people living in the area per one **Hospital**
- [6 ] Number of people living in the area per one **Pharmacy**
- [7 ] Number of people living in the area per one **Nursery Home**
- [8 ] Number of people living in the area per one **Dental Practice**
- [9 ] Number of people living in the area per one **General Practice**
- [10] Number of people living in the area per one **Health Centre**

Abovementioned statistics depicts every thirty-one county in Ireland on a high level. These are the most insightful factors which potentially can influence decision in which area should we focus more with seeking for a good place for the new hospital. In terms of frequency of people living in the area, Dublin City is the most crowded place before Cork City and Cork County. If we have a look at demographic elderly factor calculated as the ratio between Elderly people (Age 60+) and the whole population, Mayo has got the highest penetration (24%) of people in this age. The youngest penetration is representing in Fingal (13%).

On the other side we have got Demographic Load Factor that represents the ratio between people who are not in production age (younger than eighteen and older than sixty years old) and those who are in production age (older than eighteen and younger than fifty-nine years old). Dublin City with the score equal to 57% seems to have the best proportion in terms of potential workforce. Cork City is in the second position (64%). Also Fingal, Kildare and Dún Laoghaire-Rathdown could be potential interesting places for future investment with the good Demographic Load Factor score (74%, 78% and 79% respectively). The forth metric is Health Performance calculated as the ratio between those whose health is good and very good related to the whole population. Kildare, Meath and Cork County has got the healthy population across the bas where the lowest score is in Cork City and Longford. Score for Dublin City is equal to eighty-three per cent. Number of people assigned to one Hospital has peak in Kildare where there is only one hospital onsite. It's worth also to see the location of those cities visually to understand the trend that people from one county usually are travelling to another if needed. The fact that Dublin City has got the lowest number (50 413 people) is represented by approximately eleven hospitals. It's extremely important to keep in mind that every hospital is different, counties are located in different part of the country. There is different network of roads, different affluence index and each hospital can have different specialization and staff. People will be commuting from one are to another to get the support that they need and it's a natural trend as well as many other factors that can be discover locally and that are not initially visible in the data.

The lowest penetration of the people assigned to one pharmacy is in Cork City. The highest one is in Roscommon. Laois has got the highest ratio of people assigned to one nursery home (14,116). Dental Practices that are so crowded all the time are constantly busy in Laois with the 8,470 assigned to only one dental practice. Even in Dublin City the number of 1880 peoples – that is a lot. It is 156 people per month and approximately five per day per one Dental Practice. County Meath and Wexford have the highest penetration rate of General Practices while Cork City, Galway City and Dublin City has the lowest one. It doesn't mean that there is enough number of GP's in those counties. Some of those people may live in one county and commute to another for work and treatment. Finally, the last factor shows the Health Centre utilization, which is the busiest in County Meath and Kidare. Dublin City with the 8,944 people is in the 3<sup>rd</sup> position. In that way there is possible to identify clusters or ranking of areas to focus on with further analysis. In this research a strategic decision has been made, that due to population living in that area, the number of homeless, affluence indexation elderly/load factorization, penetration of abovementioned point of interests and logistically good location, the special attention will be paid to Dublin City as the key area of research.

## COUNTY MACROECONOMIC VIEW AND INSIGHTS

Dublin City is one of the most popular and well-known cities in Ireland and the whole European Union. It is also one of the most expensive and complicated cities in that part of world. In order to understand the trends and demographic it is worth to deep dive into the high level statistics derived from ABTs. In CRISP-DM this phase is called modeling.

The key actionable metrics that I derived are grouped into sixteen tables listed below:

- Key demographic metrics
- Population by age
- Population by marital status
- Population by nationality
- Size of family
- Families by family cycle
- Number of children born by Females
- Private households by type of occupancy

Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital, continued

- Socio-economic Status
- Population by social class
- Affluence Index
- Valid car driving license holders
- General health status
- Persons at work by industry
- Main Health POI institutions stats
- New building connections

From technical side, it's a V layer that represents View in MVC architecture framework. From the source code level, it works based on two steps. First, is to assign appropriate ranges (can be defined by business) and next apply that style to appropriate metric presented mellow:

```
PROC FORMAT;    value t1_Metric    /*PERCENTAGES*/
                low-0.4          = '#FF3030'
                0.4<-0.8        = '#FFFF00'
                0.8<-high       = '#7FFF00';run;
PROC PRINT DATA = CENSUS.table1_DEMOGRAPHIC_key_metrics noobs;
  var Metric Territory Ireland;
  var Territory_Share / STYLE=[BACKGROUND=t1_Metric. FONT_WEIGHT=BOLD];
run;
```

Below presented an output in SAS results window with and without applying the prepared style:

Metric	Territory	Ireland	Territory_Share
Population Size	554,554	4,761,865	11.6%
Number of Males	272,270	2,354,428	11.6%
Number of Females	282,284	2,407,437	11.7%
Number of Families	125,200	1,218,370	10.3%
Number of Households	211,591	1,697,665	12.5%
Demographic Elderly Factor *	17	591	12.5%
Demographic Load Factor **	57	2,644	2.2%
Health Performance [%] ***	83	87	95.1%

**Output 1. Country Macroeconomic View**

Metric	Territory	Ireland	Territory_Share
Population Size	554,554	4,761,865	11.6%
Number of Males	272,270	2,354,428	11.6%
Number of Females	282,284	2,407,437	11.7%
Number of Families	125,200	1,218,370	10.3%
Number of Households	211,591	1,697,665	12.5%
Demographic Elderly Factor *	17	591	12.5%
Demographic Load Factor **	57	2,644	2.2%
Health Performance [%] ***	83	87	95.1%

**Table 4. Key demographic metrics**

The intention was to incorporate my insights with the MIS (Management Information System) reports and dashboards for CxO business people. Before sharing the insights with C-Suite level Business Decision Makers, I prepared also the framework that uses SAS ODS (Output Delivery System). Thanks to the prepared and applied style visible above on the right side, L&F (Look & Feel) of the report was more pleasant, intuitive and those insights were presented in more user-friendly, easy to understand way. The ODS has got disabled STARTPAGE (STARTPAGE=NO), escape character set as '^' (ODS ESCAPECHAR='^') and styles applied to the title and footnote. Also, it's important to remember about good habit and properly close the ODS stream at the end of code via ODS \_ALL\_ CLOSE or ODS PDF CLOSE instruction.

From pure Analytical point of view, across 4.7 million of people there are 554554 people living in Dublin County which is approximately 11.6% of territory share so Dublin City is under indexed in that field. The same about most of the rest KPIs, like number of males and females that took respectively 11.6% and 11.7%. Almost 10% of all families in Ireland live in Dublin City where 12.5% of all Irish households are located. Demographic Elderly Factor presented in macroeconomic country level (17%) related the whole country (as the denominator) is equal to 12.5% which is under indexed. It means that in Dublin City are living a lot of young people who are in production age (Students, young families etc.). Demographic Load Factor presented in macroeconomic country level (57%) related the whole country (denominator) is equal to 2.2% which is very low and under indexed. It means that 2.2% of all the population of Ireland that is not in production age is living in Dublin City. Knowing the main statistics we can dig deeper to understand the age of population living in Dublin City:

Age_Group	Territory	Ireland	Index
1) Age 19	2.0%	2.0%	100
2) Age 20-24	10.0%	10.0%	100
3) Age 25-29	14.0%	8.0%	175
4) Age 30-34	13.0%	10.0%	130
5) Age 35-39	11.0%	11.0%	100
6) Age 40-44	8.0%	10.0%	80
7) Age 45-49	8.0%	9.0%	89
8) Age 50-54	7.0%	9.0%	78
9) Age 55-59	6.0%	8.0%	75
10) Age 60-64	5.0%	7.0%	71
11) Age 65-69	5.0%	6.0%	83
12) Age 70-74	4.0%	5.0%	80
13) Age 75-79	3.0%	3.0%	100
14) Age 80-84	2.0%	2.0%	100
TOTAL	98.0%	100%	.

**Table 5. Population by age**

Table above presents different age brackets. We can see that in our area there is living the same percentage of people in Age 19-24, Age 35-39, Age 75-84 like in average of Ireland. We as a county are over indexing in percentage of population of Age 25-29 (by 75%) and of population of Age 30-34 (by 30%). On the other site there is 17% people less in age 65-69 than in average of the country and 29% less of the percentage of people in age 60-64 than in average of the country.

Status	Territory	Ireland	Index
1) Single	60.0%	54.0%	111
2) Married (incl. same sex civil partnership)	31.0%	38.0%	82
3) Separated	2.0%	2.0%	100
4) Divorced	2.0%	2.0%	100
5) Widowed	4.0%	4.0%	100
TOTAL	99.0%	100%	.

**Table 6. Population by marital status**

Size_of_Family	Territory	Ireland	Index
1) 2 persons (No. of families)	49%	40%	123
2) 3 persons (No. of families)	23%	23%	100
3) 4 persons (No. of families)	18%	22%	82
4) 5 persons (No. of families)	8%	11%	73
5) 6 or more persons (No. of families)	3%	5%	60
TOTAL	101%	101%	.

**Table 7. Population by the size of family**

The next insight is about the number of enumerated residents by marital status. Almost 11% more of people living in the area of Dublin City is single that average across the whole country. It confirms the trend that was mentioned above, that there are a lot of young generation representatives who are living here in Dublin City. That brings a lot of opportunities and challenges due to their expectations and characteristic of that segment of population. Majority of people living in Dublin are Irish (77%). The next nationality (PROC FREQ are British and on 3<sup>rd</sup> place there are Polish. An interesting trend represents the nationals outside European Union and the Rest of the world (100% more people like an Irish average).

When it comes to the family, the typical model is the family that consists of 3 persons (two parents with one child) and more often only two persons. Those two persons, who are over indexing by 23% can represents ever-married cohabiting couples and/or lone mothers (who are both never-married and divorced or separated). The percentage in the Territory and Ireland level equal to 101% was calculated with rounding up (that's why it's >100%).



Family_Cycle	Territory	Ireland	Index
1) Pre-family (No. of families)	18%	9%	200
2) Empty nest (No. of families)	8%	10%	80
3) Retired (No. of families)	9%	10%	90
4) Pre-school (No. of families)	11%	10%	110
5) Early school (No. of families)	10%	12%	83
6) Pre-adolescent (No. of families)	9%	11%	82
7) Adolescent (No. of families)	10%	12%	83
8) Adult (No. of families)	26%	25%	104
TOTAL	101%	99%	.

**Table 8. Population by family cycle**

Number_of_Children_Born	Territory	Ireland	Index
1) 0 children born	46%	31%	148
2) 1 child born	13%	13%	100
3) 2 children born	17%	22%	77
4) 3 children born	12%	17%	71
5) 4 or more children born	13%	18%	72
TOTAL	101%	101%	.

**Table 9. Population by children born by Females**

By using advance statistical techniques CENSUS data in Ireland majority of families in Ireland are outside the 'traditional model'. Understanding the modern family is critical if we want to design and deliver the high quality health services (hospital). The findings from the study shows, that the in terms of families cycle (life stage), almost 100% more of the families in Dublin City territory are in Pre-Family life stage. The number of retired families is lower than average across the country due to presence of younger population.

Data shows clearly that family structure in Ireland moving beyond the traditional model. There are 48% more females in Dublin City than average in Ireland, who has no children. Younger people have got different experience and different needs that documented above data.

Apart from the people's life style, also market is rapidly changing bringing the new challenges and opportunities. As the market property prices are still increasing, more people decide to rent the household from Local Authority or private landlord. Dublin is great example of the cities where (like in San Francisco) the shortage of housing grew year by year, rising to a deficit of homes of more than 50,000 by 2015<sup>5</sup>.

Economic_Status	Territory	Ireland	Index
1) At work	56%	53%	106
2) Looking for first regular job	1%	1%	100
3) Unemployed having lost or given up previous job	7%	7%	100
4) Student	11%	11%	100
5) Looking after home/family	6%	8%	75
6) Retired	14%	15%	93
7) Unable to work due to permanent sickness or disability	4%	4%	100
8) Other	1%	0%	.
TOTAL	100%	99%	.

**Table 10. Socio-economic Status**

<sup>5</sup> Source: [www.politico.eu/article/ireland-housing-shortage-puts-strain-on-revival/](http://www.politico.eu/article/ireland-housing-shortage-puts-strain-on-revival/)

Advanced Analytics approach to explore CENSUS data and know the population better in order to determine the best opportunities to open the new Hospital, continued

Majority of people are employed (6% higher than an average) and 4% of the whole Dublin City population is unable to work due to permanent sickness or disability. Not only for those target group, might the investment in hospital be beneficial in terms of future health and wellbeing.

Social_Class	Territory	Ireland	Index
1) Professional workers	10%	8%	125
2) Managerial and technical	27%	28%	96
3) Non-manual	16%	18%	89
4) Skilled manual	11%	14%	79
5) Semi-skilled	11%	14%	79
6) Unskilled	4%	4%	100
7) All others gainfully occupied and unknown	23%	18%	128
TOTAL	102%	104%	.

**Table 11. Population by Social Class**

General_Health	Territory	Ireland	Index
1) Very good	56%	59%	95
2) Good	27%	28%	96
3) Fair	8%	8%	100
4) Bad	2%	1%	200
5) Very bad	0%	0%	.
6) Not stated	7%	3%	233
TOTAL	100%	99%	.

**Table 12. General Health Status**

Majority of Dubliners represents a group of professional workers. It also reflects affluence index across the whole population calculated as percentage of professional workers and managerial and technical workers. Also the study showed an interesting insight, that only 25% of the Dublin society has got a car driving license. Most of them are facing a problem with traffic and commuting day by day by public transport so that's the reason of such low number of the driving license holders. Additionally, within those 630156 people there are also a people who are too young to have the driving license.

People generally happy with the level of health status they have. But those are the young people who are do not pay attention on many factors like drugs, cigarettes; alcohol etc. 100% people more than in average of Ireland claimed that their health is bad. That's another justification of focusing on Dublin City in terms of the potential localization of the new hospital. It is not only for taking care of those with bad health but also to make the young with very good, good and fair health status secure in the future.

Industry	Territory	Ireland	Index
1) Agriculture, forestry and fishing	0%	4%	0
2) Building and construction	3%	5%	60
3) Manufacturing industries	5%	11%	45
4) Commerce and trade	27%	24%	113
5) Transport and communications	13%	9%	144
6) Public administration	5%	5%	100
7) Professional services	23%	24%	96
8) Other	23%	18%	128
TOTAL	99%	100%	.

**Table 13. Persons at work by industry**

Majority of Dubliners are working in Commerce and trade, Transport and communication and other industries. This shows also that the power of affluence index that represents percentage of professional workers and managerial and technical workers. Those young people might be working in big corporations.

The next two tables below presents main health Points of Interest. Regarding to the number of airports, it's good to know that in advance (as well as the proximity to the closest airport) in case of patient transportation). The number of Dental Practices and GP's will show how the number of our potential competitors look. The same case is with registered Health Centers, Hospitals, Pharmacies and Nurse Homes.

Object	Territory	Ireland	Index
1) Number of present Airports	1	8	13
2) Number of registered Dental Practices	295	1,186	25
3) Number of registered General Practices	638	2,335	27
4) Number of registered Health Centres	62	835	7
5) Number of registered Hospitals	11	38	29
6) Number of registered Pharmacies	460	2,025	23
7) Number of registered Nurse Homes	109	693	16

Metric	Territory	Ireland	Index
1) ESB Connection Summary (2014)	914	10,929	8
2) ESB Connection Summary (2015)	667	12,583	5
3) ESB Connection Summary (2016)	731	9,095	8

**Table 14. Main Health Point of Interest (POI)**

**Table 15. New building connections**

As the Dublin's housing crisis is progressing, making Dublin challenging location for business, it is worth to know how many new building connection have been made in last years. These data are based on the number of new dwellings connected by the ESB to the electricity supply but exclude conversions and demountable. This dataset may also include units at some unfinished housing developments which have been reactivated in line with the recovery in the housing market. This dataset may also include units at some unfinished housing developments which have been reactivated in line with the recovery in the housing market. They may not accord precisely with local authority boundaries but that's the quality of data that we need to accept in this case.

## CASE STUDY: REAL-WORLD EXAMPLE OF PATIENT-CENTERED HEALTHCARE

It is always good to see how this timeless knowledge and insights can be applicable to the real life. One of the best world examples is the New Children's Hospital (NCH), its satellite centers and tri-location. The critical decision about the location of the new hospital was supported by the proximity data, information about the range of national specialties, education, culture and infrastructure in that area. One of the factor that decision makers took into consideration was also the number of beds and duplication or even triplication of some clinical services within a 14 kilometers distance across many hospitals around the city.

According to the NCH report<sup>6</sup>, by integrating the three existing Dublin children's hospitals into one single large hospital and co-locating with a major adult teaching hospital on the campus at St James's Hospital, the new children's hospital will operate the system proven to deliver better clinical outcomes in most leading children's hospitals across the world, that is scale and critical mass combined with adult co-location. For major public tertiary children's hospitals scale (i.e. the highest volume of the most complex cases), critical mass and tri-location demonstrably improve clinical outcomes and save more children's lives by optimizing the utilization and effectiveness of multiple clinical teams of specialists and subspecialists and expensive equipment.



**Display 2. Children's Hospital Group Case Study**

<sup>6</sup> Children's Hospital Group Case Study

<http://www.nchplanning.ie/wp-content/uploads/2015/07/3-Clinical-Case-for-New-Childrens-Hospital.pdf>

## CONCLUSION:

As stated at the beginning, selecting the best place for the new hospital is absolutely always a big challenge. Rapidly changing market conditions, demanding society, market seasonality and sensitivity, political factors (like Brexit). There are multitude scenarios that you need to consider (traffic jams, ambulance call outs frequency etc.) and there will be always some level of uncertainty. It's impossible to mitigate all of the risks but in order to minimize the ratio of wrong decision, it is better to rely on the hard data and facts rather than assumptions. Intuition driven approach is not the best way to tackle this kind of problem.

There are no golden rules and crystal ball that helps to made an appropriate decision. It is rather (and that's the reason why I am pointing that out) a universal, applicable framework that you can adjust to your organization purpose(s) and needs. That is why I proposed to focus and to rely on hard data and build fact base like for your Planning Process (PP), presented in this research that shows the macro and micro economic conditions. You can also incorporate the PESTLE analysis into this research and use geospatial technology for gap analysis and territory planning.

In conjunction with knowledge and information that your organization owns, the final go-to-market or not-go-to-market decision related to particular location will be more suitable and more accurate. The goal of this research was to show Advanced Analytics approach in real life example. My study shows how to explore CENSUS and open data and know the population and market characteristic better in order to determine the best opportunities to open the new Hospital. Those actionable and universal applicable insights are presented from a high level, as this is a goal of this research. Obtaining this "helicopter view" can help ensure the final, rational data-driven decision is made with full consideration of the chances, implications, population needs and potential opportunities – it's about saving people life.

## RECOMMENDED READING:

- Republic of Ireland: [https://en.wikipedia.org/wiki/Republic\\_of\\_Ireland](https://en.wikipedia.org/wiki/Republic_of_Ireland)
- The Central Statistics Office: [www.cso.ie/en/census/](http://www.cso.ie/en/census/)
- DAO (**D**ata **A**ccess **O**bject) software design pattern: [https://en.wikipedia.org/wiki/Data\\_access\\_object](https://en.wikipedia.org/wiki/Data_access_object)
- MVC (**M**odel–**V**iew–**C**ontroller) software design pattern: <https://en.wikipedia.org/wiki/Model–view–controller>
- Irish Eircode: <https://www.dccae.gov.ie/en-ie/communications/topics/Postcodes/Pages/default.aspx>
- Irish Eircode: <https://www.eircode.ie/getting-an-eircode>

## ACKNOWLEDGMENTS

I would like to cordially thank Marta Wasiak for the openness to literally every extremely ambitious and interesting idea that I have had, for her guidance and encouragement. Thank you for an inspiration, sharing your wisdom and passion to Advance Analytics with me.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Krystian Matusz  
[matuszkrystian@gmail.com](mailto:matuszkrystian@gmail.com)  
[www.linkedin.com/in/krystianmatusz/](http://www.linkedin.com/in/krystianmatusz/)