



Abide Data Analysis

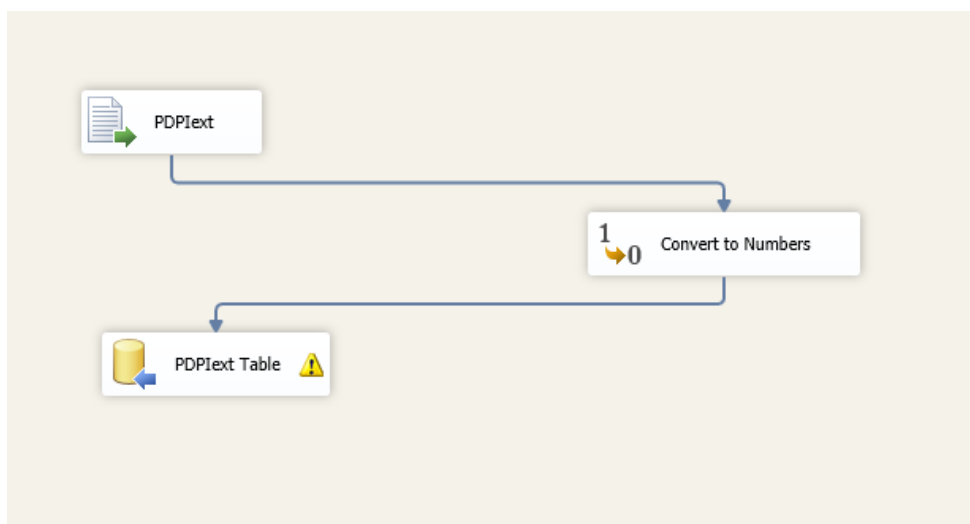
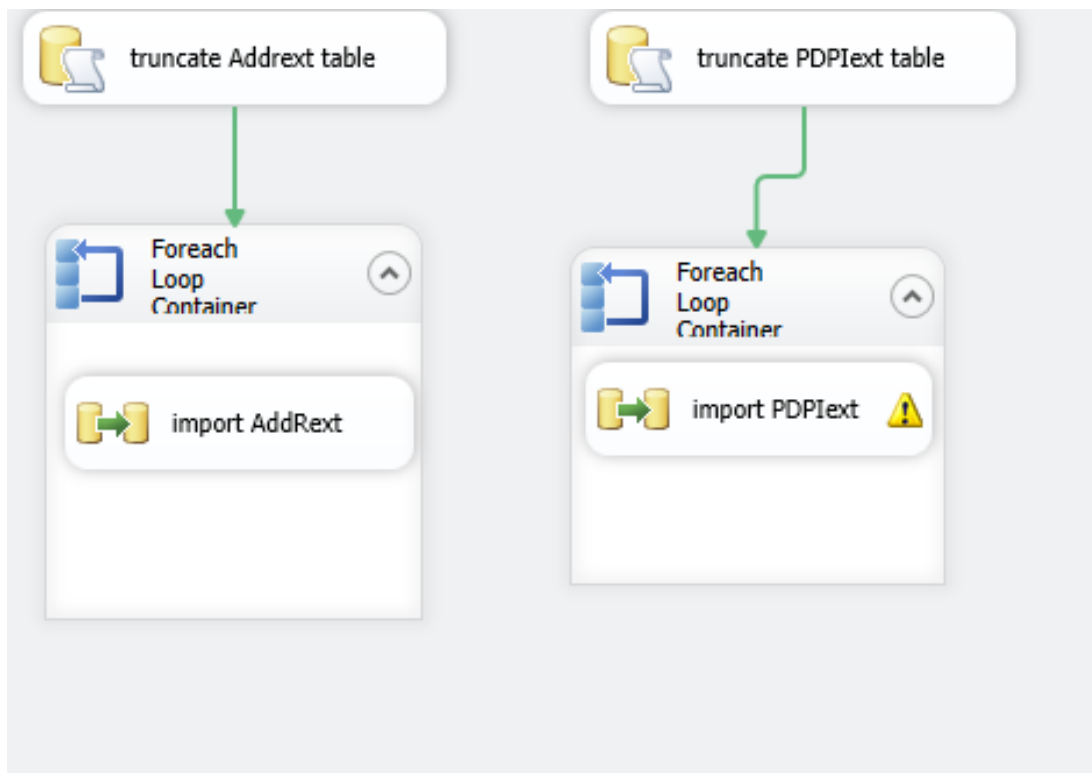
Abide Data Project

Table of Contents

Introduction	2
Question 1.....	3
Question 2.....	4
Question 3.....	5
Question 4.....	7
Question 5.....	8
Conclusions	11
References	12
Appendix	13
1.Tables	13
Question1.....	13
Question 2.....	13
Question 3.....	13
Question 4.....	14
Question 5.....	14
2.Glossary.....	14

Introduction

Before answering any of the questions, I had to use SSIS (2012) to load the tables to the database. For initial analysis, Excel 2013 is not able to open the big files.



The England postcode table was loaded from access to SQL database. The source of the table is in appendix. The AddRext table without column names was named like this;

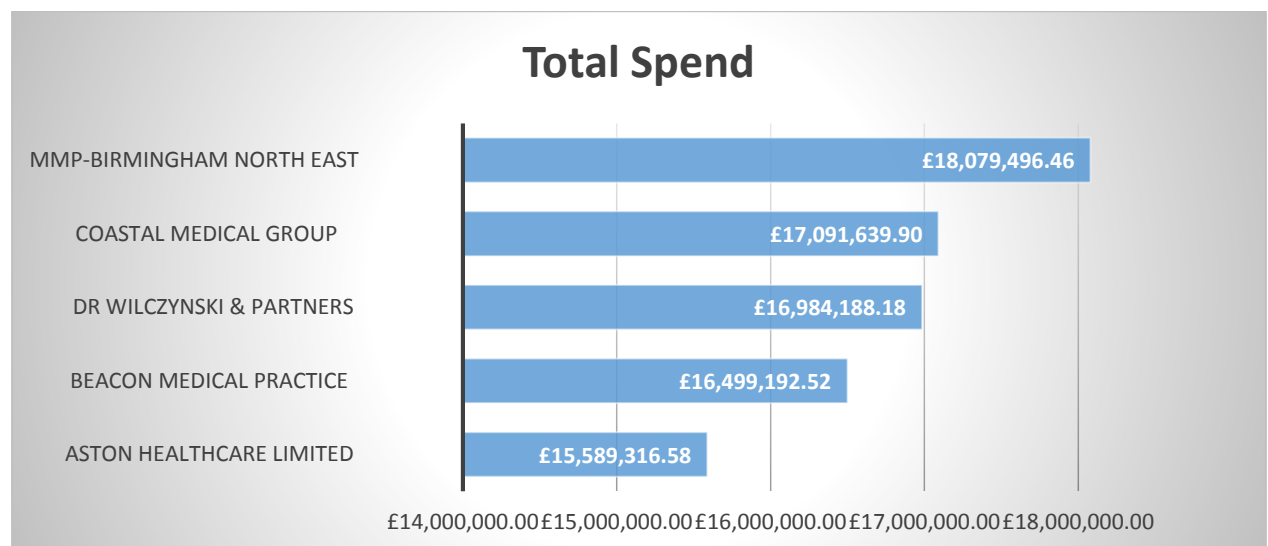
PERIOD, PRACTICE_CODE, NAME, ADDRESS1, ADDRESS2, ADDRESS3, ADDRESS4, POSTCODE

Question 1

USE AbideDB;

```
/*  
*****  
Calculate top 5 post Codes in terms of Total spending on Prescriptions  
*****  
*/
```

```
SELECT TOP(5) WITH TIES  
    Name,  
    Practice,  
    PostCode,  
    SUM (ISNULL((Act_Cost),0)) AS Spend  
FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =  
AddRext.PRACTICE_CODE  
GROUP BY Name, Practice, PostCode  
ORDER BY Spend DESC
```



Question 2

USE AbideDB;

```
/*  
Calculate Top 5 Practices in terms of Total spending on Diclofenac Prescriptions  
*/
```

```
SELECT TOP(5) WITH TIES  
    Practice,  
    NAME,  
    SUM(ISNULL(CAST(ITEMS AS INTEGER),0)) AS PRESCRIPTIONS  
  
FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =  
AddRext.PRACTICE_CODE  
  
WHERE BNF_NAME LIKE '%Diclofenac%'  
GROUP BY NAME, Practice  
ORDER BY PRESCRIPTIONS DESC
```

Results using SSRS

DiclofenacReport.rdl [Design] X

Design Preview

1 of 1 100% Find

Practice Code	Practice Name	No of Prescriptions
C83019	BEACON MEDICAL PRACTICE	18006
A82044	FELLVIEW HEALTHCARE LTD	15288
C84024	NEWGATE MEDICAL GROUP	15222
M85063	MMP-BIRMINGHAM NORTH EAST	13950
N83028	ASTON HEALTHCARE LIMITED	13098

Question 3

```
USE AbideDB;

/*****

Bottom 5 PCT who spent the least & Mean (AvgSpend), Deviations, ZScore (deviation
from the Mean)

*****/

SELECT TOP(5)

    Name,

        --****      Spend per Practice      ****

    SUM(ISNULL(Act_Cost ,0)) AS Spend,

        --****      Average Spend in the country      ****

    (SELECT AVG(Spend) FROM (SELECT Name, SUM(ISNULL(Act_Cost ,0)) AS Spend
    FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
    AddRext.PRACTICE_CODE GROUP BY Name)c ) AS AVGSpend,

        --****      Standard Deviation of Spend in the country      ****

    (SELECT STDEVP(Spend) FROM (SELECT Name, SUM(ISNULL(Act_Cost ,0)) AS Spend
    FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
    AddRext.PRACTICE_CODE GROUP BY Name)d ) AS STDevSpend,

        --****      ZScore: Number of Standard Deviations from the Country Mean      ****

    (SUM(ISNULL(Act_Cost ,0)) -
    (SELECT AVG(Spend) FROM (SELECT Name, SUM(ISNULL(Act_Cost ,0)) AS Spend
    FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
    AddRext.PRACTICE_CODE
    GROUP BY Name)c ) )/
    (SELECT STDEVP(Spend) FROM (SELECT Name, SUM(ISNULL(Act_Cost ,0)) AS Spend
    FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
    AddRext.PRACTICE_CODE GROUP BY Name)d ) AS ZScore

FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
AddRext.PRACTICE_CODE

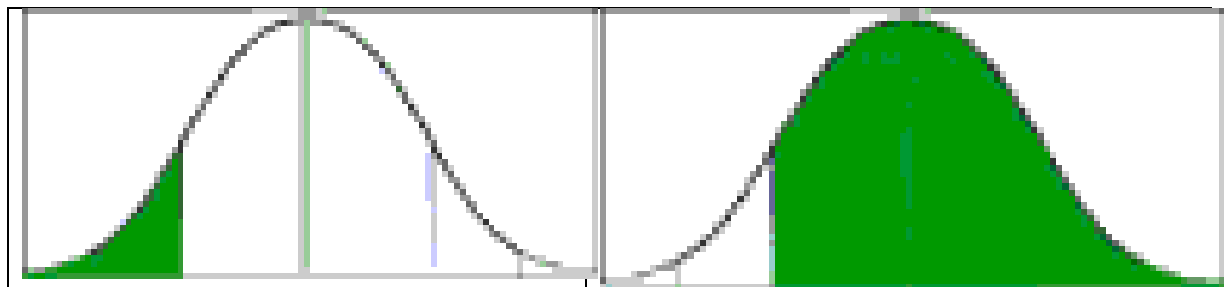
GROUP BY Name
ORDER BY Spend ASC
```

A Z score allows you to take any given sample within a set of data and to determine how many standard deviations above or below the mean it is. To find the Z score of a sample, you'll need to find the standard deviation and mean of a set of data, find the difference between that sample and the mean, and divide it by the standard deviation.

Results

Name	Spend	AVGSpend	STDevSpend	ZScore
MEADOWBROOK/LDP	0.40	2250102.24016	2,314,482.39371555	-0.972183606196203
THE TIWARI AND MALIK PRACTICE	0.52	2250102.24016	2,314,482.39371555	-0.972183554348757
HMP BIRMINGHAM	0.54	2250102.24016	2,314,482.39371555	-0.972183545707516
SINGLE POINT OF ACCESS (SPA)	0.62	2250102.24016	2,314,482.39371555	-0.972183511142553
HMP GRENDON	0.69	2250102.24016	2,314,482.39371555	-0.972183480898209

This can be useful if you want to compare your test results to a larger population. Selecting one side, and looking it up on the Z Table 16.54%. This means that if you look up your score on the bell curve, that 83.45% of data would have scored on or above the Z-value.



Question 4

```
USE AbideDB;
GO
```

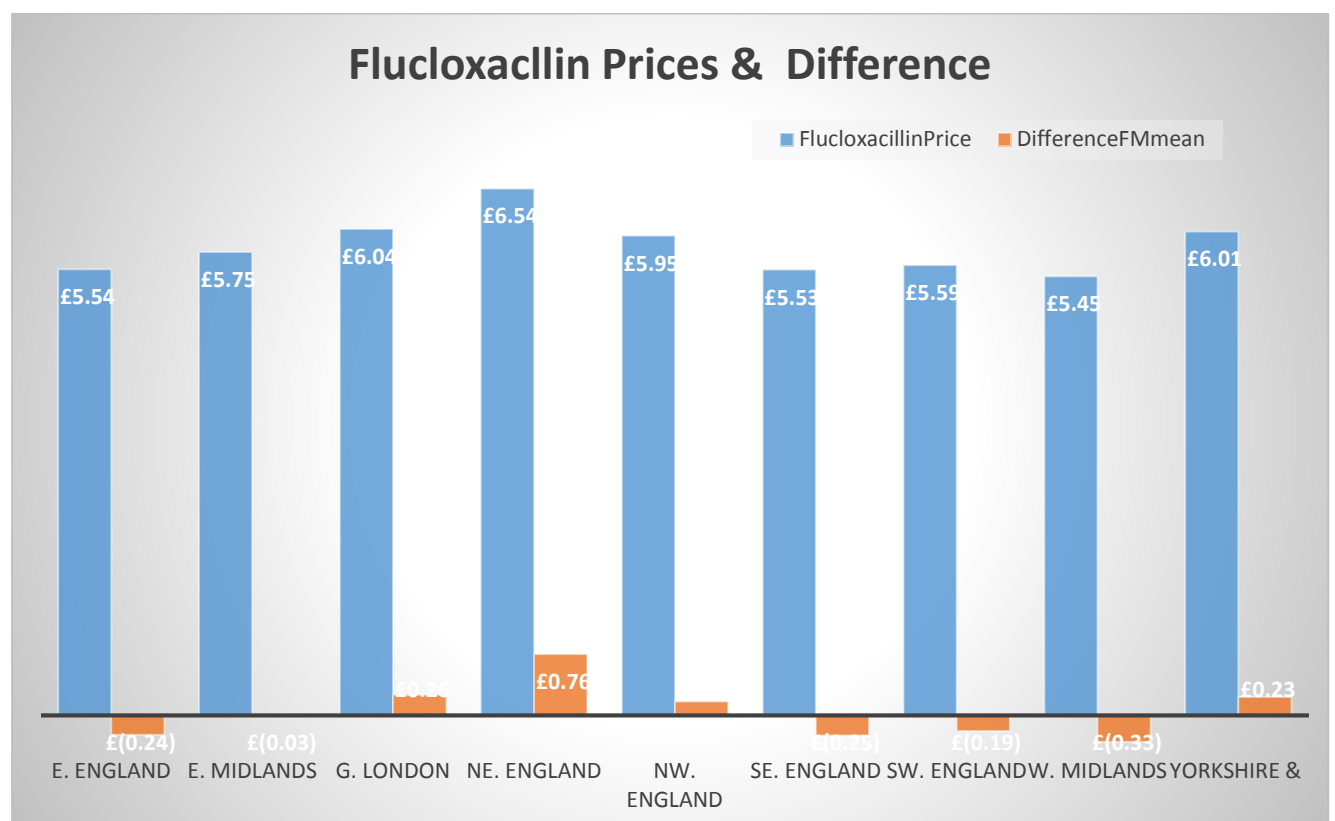
```
/******
Question 4: Average price of Flucloxacillin & Variation from the mean
******/
```

```
*****/
```

```
SELECT Region,
        AVG(Act_Cost/Items) AS FlucloxacillinPrice,
        (AVG(Act_Cost/Items) - (SELECT AVG(Act_Cost/Items )
                                FROM PDPIext
                                LEFT OUTER JOIN AddRext ON PRACTICE = PRACTICE_CODE
                                LEFT OUTER JOIN EnglandPostCodes ON LEFT(AddRext.Postcode, 4) =
LEFT(EnglandPostCodes.Postcode,4)
                                WHERE [BNF_NAME] LIKE '%Flucloxacillin Sodium%' OR [BNF_NAME]
LIKE '%Flucloxacillin Magnesium%'
                                )) AS DifferenceFMmean
```

```
FROM PDPIext
LEFT OUTER JOIN AddRext ON PRACTICE = PRACTICE_CODE
LEFT OUTER JOIN [EnglandPostCodes] ON LEFT(AddRext.Postcode, 4) =
LEFT(EnglandPostCodes.Postcode,4)
```

```
WHERE [BNF_NAME] LIKE '%Flucloxacillin Sodium%' OR [BNF_NAME] LIKE '%Flucloxacillin
Magnesium%'
GROUP BY Region
```



Question 5

Using the above tables draw Boxplot and frequency diagrams for prescription spending in the country.

```
USE AbideDB;
```

```
/****** Script for Prescription Spending Table *****/
```

```
SELECT * INTO GPSPendTable1 FROM (
SELECT
    Name,
    Practice,
    PostCode,
    SUM(ISNULL(Act_Cost,0)) AS Spend
FROM PDPIext LEFT OUTER JOIN AddRext ON PDPIext.PRACTICE =
AddRext.PRACTICE_CODE
GROUP BY Name, Practice, PostCode
ORDER BY Spend DESC
)n
```

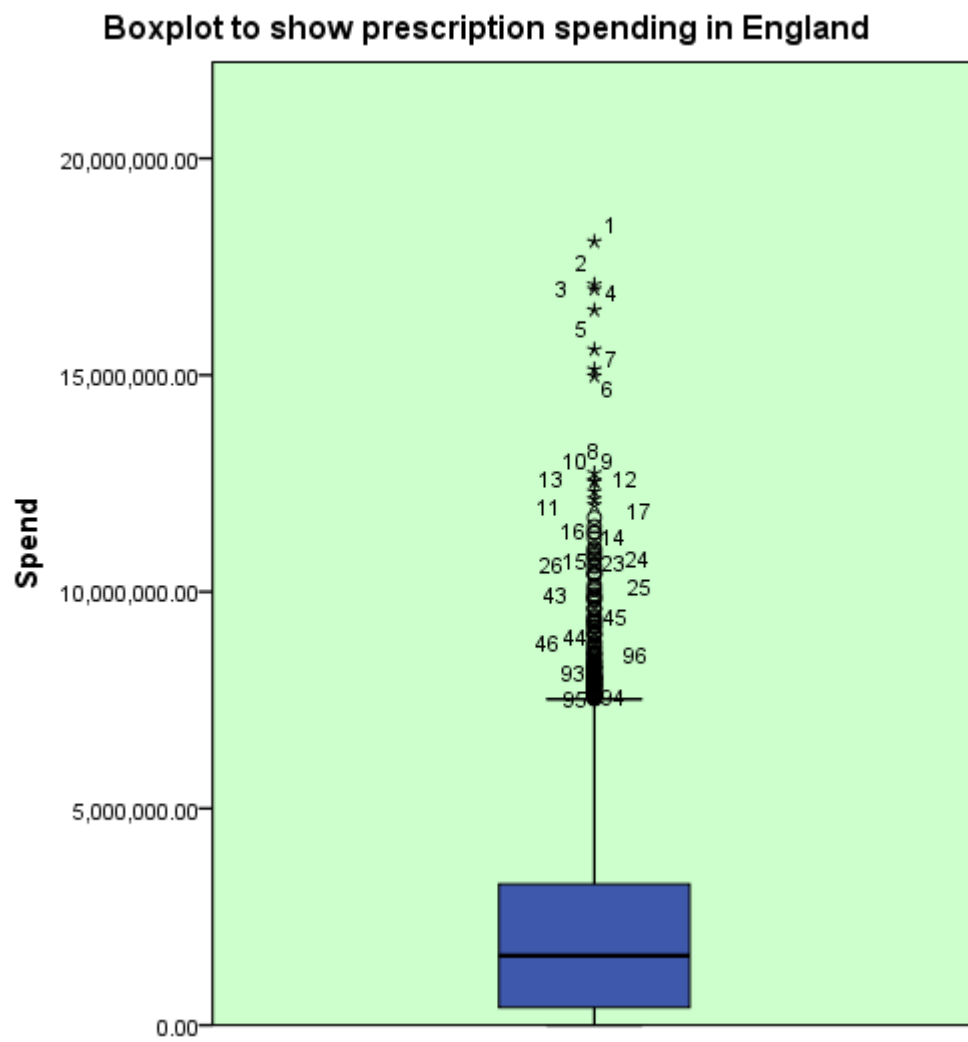
```
/****** Script for Prescription Spending *****/
```

```
SELECT [Name],
       [Practice],
       [PostCode],
       [Spend]
FROM [AbideDB].[dbo].[GPSPendTable]
```

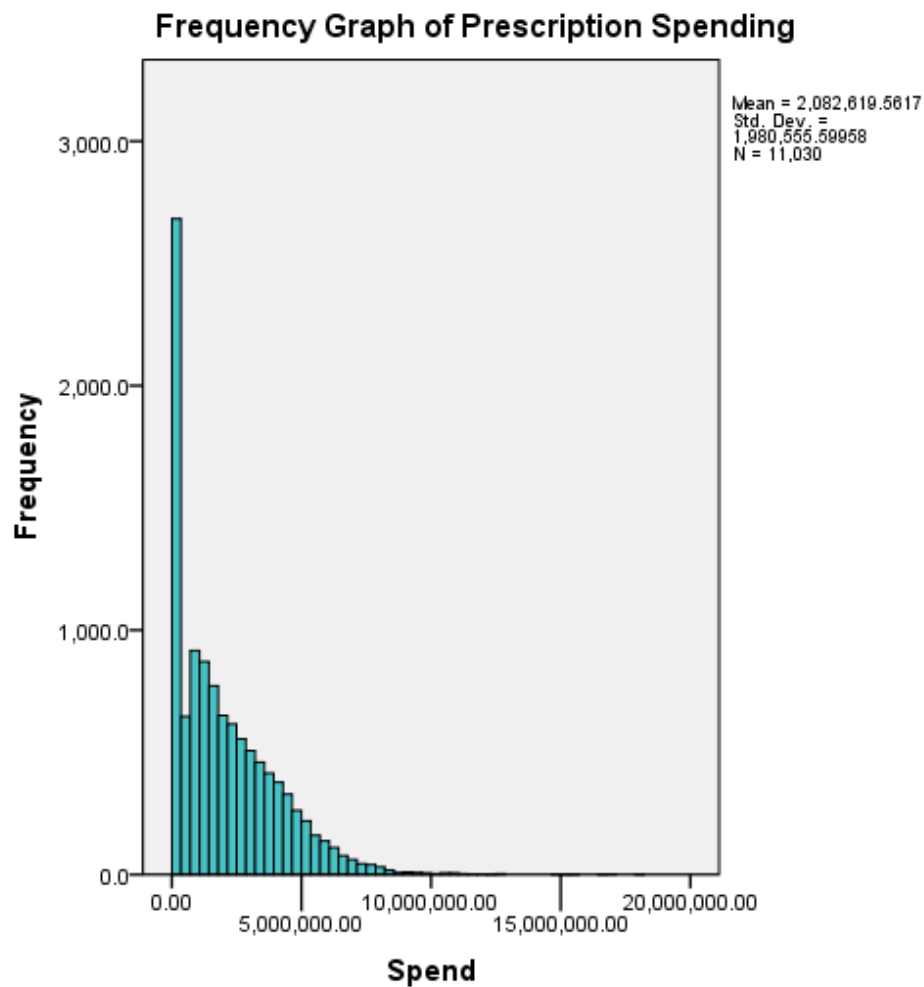
The following Statistics were calculated using IBM SPSS. You can get the same values if you use T-SQL.

Statistics		
Spend		
N	Valid	11030
	Missing	0
Mean		2,082,619.5617
Std. Error of Mean		18,858.15854
Median		1,603,818.6500
Mode		1.07
Std. Deviation		1,980,555.5995
		8
Variance		3.923E12
Skewness		1.273
Std. Error of Skewness		.023
Kurtosis		2.715
Std. Error of Kurtosis		.047
Range		18,079,495.60
Minimum		0.40
Maximum		18,079,496.00
Sum		2.30E10
Percentiles	25	408,471.7100
	50	1,603,818.6500
	75	3,254,989.1500

From the above table and SQL tables the box plot can now be drawn.



The box plot for Spending shows stars, indicating that there are a lot of outliers in this distribution.



The graph shows positive skewness. The right tail is longer, and the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed.

Conclusions

1. The addresses are not complete, although they seem to have the required postcodes. From the post codes, one has then to find out where they are.
2. England is not the only area covered, I have found few addresses in Northern Ireland, Scotland and Wales.
3. With the ongoing post code changes, incomplete information, postcode information in different places and some information chargeable, it is very difficult to come up with an exact picture of geographically medical data.

References

Glossary of Terms for Prescribing Reports

<http://www.nhsbsa.nhs.uk/PrescriptionServices/3197.aspx>

Postcode CSV downloads

<http://www.doogal.co.uk/PostcodeDownloads.php>

Information about UK towns, counties & postcodes

<http://www.townscountiespostcodes.co.uk/postcodes-in-england/>

UK regions

<http://uk-regions.page.co.uk/>

<http://www.fieldstar.org/post-codes-regions-list>

Data

<http://data.gov.uk/dataset/gp-practice-prescribing-data>

Appendix

1.Tables

Question1

Name	PostCode	Spend
ASTON HEALTHCARE LIMITED	L36 0UB	£ 15,589,316.58
BEACON MEDICAL PRACTICE	PE25 2RN	£ 16,499,192.52
DR WILCZYNSKI & PARTNERS	NN17 2UR	£ 16,984,188.18
COASTAL MEDICAL GROUP	LA4 5LY	£ 17,091,639.90
MMP-BIRMINGHAM NORTH EAST	B24 8NT	£ 18,079,496.46

Question 2

Practice	NAME	PRESCRIPTIONS
C83019	BEACON MEDICAL PRACTICE	18006
A82044	FELLVIEW HEALTHCARE LTD	15288
C84024	NEWGATE MEDICAL GROUP	15222
M85063	MMP-BIRMINGHAM NORTH EAST	13950
N83028	ASTON HEALTHCARE LIMITED	13098

Question 3

Name	Spend	AVGSpend	STDevSpend	ZScore
MEADOWBROOK/LDP	0.40	2250102.24016	2,314,482.39371555	-0.972183606196203
THE TIWARI AND MALIK PRACTICE	0.52	2250102.24016	2,314,482.39371555	-0.972183554348757
HMP BIRMINGHAM	0.54	2250102.24016	2,314,482.39371555	-0.972183545707516
SINGLE POINT OF ACCESS (SPA)	0.62	2250102.24016	2,314,482.39371555	-0.972183511142553
HMP GRENDON	0.69	2250102.24016	2,314,482.39371555	-0.972183480898209

Question 4

Region	FlucloxacillinPrice	DifferenceFMmean
E. England	£ 5.54	-£ 0.24
E. Midlands	£ 5.75	-£ 0.03
G. London	£ 6.04	£ 0.26
NE. England	£ 6.54	£ 0.76
NW. England	£ 5.95	£ 0.17
SE. England	£ 5.53	-£ 0.25
SW. England	£ 5.59	-£ 0.19
W. Midlands	£ 5.45	-£ 0.33
Yorkshire &	£ 6.01	£ 0.23

Question 5

The table is too big to include here.

2. Glossary

BNF Code:

The BNF Code is a 15 digit code in which the first seven digits are allocated according to the categories in the BNF and the last 8 digits represent the medicinal product, form, strength and the link to the generic equivalent product. The digits in the code represent the following information:

1 & 2	BNF Chapter
3 & 4	BNF Section
5 & 6	BNF Paragraph
7	BNF Sub-Paragraph
8 & 9	Chemical substance
10 & 11	Product
12 & 13	Strength / Formulation
14 & 15	Link to the generic equivalent product. A is used when there is no linking record.

BNF Section Name:

The primary therapeutic area based on BNF classification for which a drug is used.

Drug name:

Medicines are shown by individual preparation name, which may be proprietary or generic, followed by form and strength. The names used are often presented in an abbreviated form. The booklet Drug and Appliance Abbreviations <http://www.nhsbsa.nhs.uk/PrescriptionServices/3197.aspx> provides a list of abbreviated definitions used.

BNF Chemical Name:

PCA data is based on the therapeutic grouping used in the British National Formulary. The classification in this report is based on the September 2010 BNF (edition 60). While the therapeutic classification of drugs occasionally changes, all PCA publications reflect the classification in use at the time of publication.

The NHS Prescription Services has created pseudo BNF chapters, which are not published, for items not included in BNF chapters 1 to 15. The majority of such items are dressings and appliances, which the NHS Prescription Services has classified into four pseudo BNF chapters (20 to 23).

Class of preparation (Prep Class):

PCA data classifies drugs in four ways:

Class 1 - Drugs prescribed and available generically

Class 2 - Drugs prescribed generically but only available as a proprietary product

Class 3 - Drugs prescribed and dispensed by proprietary brand name

Class 4 - Dressings and appliances

Standard quantity unit (SQU):

This code indicates the form of the drug and the units in which quantity is measured:

Code 1 - a unit (e.g. one tablet, capsule, pack, aerosol etc)

Code 3 - millilitres

Code 6 - grammes

Code 0 - individually formulated (unit varies)

Items dispensed:

A prescription item refers to a single medicine prescribed by a doctor (or dentist/nurse/etc.) on a prescription form. This is different to quantity i.e. if salbutamol inhaler x 2 was prescribed. This is one item with a quantity of two.

Quantity (QTY):

The quantity of a drug dispensed is measured in units depending on the formulation of the product, see standard quantity unit below. Quantities are not added together across preparations because of different strengths and formulations.

Net Ingredient Cost (NIC):

Net Ingredient Cost is the basic price of the medicine, i.e. the price listed in Part VIII or IX of the Drug Tariff http://www.ppa.org.uk/ppa/edt_intro.htm or for drugs not listed in the Drug Tariff, the list price published by the manufacturer, wholesaler or supplier (See Drug Tariff Part II clause 8). This is the cost of the drug before any discounts are applied and does not include any dispensing costs or fees. It also does not include any adjustment for income obtained where a prescription charge is paid at the time the prescription is dispensed or where the patient has purchased a pre-payment certificate.

Of Which Class 2:

Shows the number of items which have been prescribed generically that have been dispensed and reimbursed using a proprietary product because a generic version of the drug is not available. The dispenser can choose any suitable proprietary drug to meet the generic order and does not necessarily have to provide information to show what he has supplied. In this situation the data is defined as class 2 but is reported in PCA against the class 3 (proprietary) form of the drug. To make this allocation NHS Prescription Services apply the following rules:

If there is only one proprietary equivalent the generic prescribing is attributed to that product. If there is more than one equivalent proprietary, the data is prorated across all the equivalent products based on the proprietary prescribing of the drug. Any remainder from the prorated calculation is added to the first proprietary equivalent as it appears on NHS Prescription Services database. This is to ensure that the amount shared between the proprietary equivalents is exactly equal to the class 2 product dispensed. The prorated calculation for each of the elements items, quantity and NIC is calculated independently therefore some preparations may show data for quantity, NIC but have no data for items because the prorated % allocation is so small.

If there is no proprietary prescribing of the equivalent proprietary products during the processing period the class 2 prescribing is prorated across all equivalent proprietary products.

Term	Definition
SHA	Strategic Health Authority
PCT	Primary Care Trust,
Practice	A GP practice is an organisation with one or more GP's acting in partnership. For the purposes of the NHS Prescription Services Information Systems, any organisation that shows information down to prescriber level can be set up as a practice. This includes Cost Centres for PCOs.
BNF codes	<p>A fifteen character code, based on the British National Formulary (BNF) classifications. The code breaks down as follows:</p> <ul style="list-style-type: none"> • Char 1 - 2 - BNF Chapter • Char 3 - 4 - BNF Section • Char 5 - 6 - BNF Paragraph • Char 7 - BNF Sub-paragraph • Char 8 - 9 - Chemical substance • Char 10 - 11 - Drug or Product • Char 12 - 13 - Strength/Formulation • Char 14 - 15 - Equivalent drug <p>For example, the codes for Lisinopril tablets 2.5mg and Zestril tablets 2.5mg are respectively 0205051L0AAAAAA and 0205051L0BBAAAA. The first two characters identify BNF Chapter 2 (cardiovascular system). The third and fourth characters identify section 5 (drugs affecting the renin-angiotensin system and some other antihypertensive drugs). The fifth and sixth characters identify paragraph 5 (drugs affecting the renin-angiotensin system). Character 7 identifies the sub-paragraph 1 (angiotensin-converting enzyme inhibitors). The eighth and ninth characters identify the substance as lisinopril. The tenth and eleventh characters identify a generic or a specific brand. The twelfth and thirteenth characters identify tablets and 2.5mg. The fourteenth and fifteenth characters link the brand to the generic.</p>
BNF Name	The primary therapeutic area based on BNF classification for which a drug is used.
Items	(Prescription Item) Items is the number of times a product appears on a prescription form
NIC	(Net Ingredient Cost) Basic price of a drug, ie the price listed in Part II Clause 8 of the Drug Tariff.
Actual Cost	The actual cost is the basic price of the drug less the National Average Discount Percentage (NADP) plus Payment for Consumables