



BIKE MS ANALYTICS

TEAM NAME: KDD DEEP LEARNERS UNCC

Analysis Focus:

“What occupations were responsible for most of BIKE MS’s fundraising?”

Motivation:

To identify occupations where people are more likely to contribute towards Multiple Sclerosis and the trend in involvement over the past years.



DATASET USED:

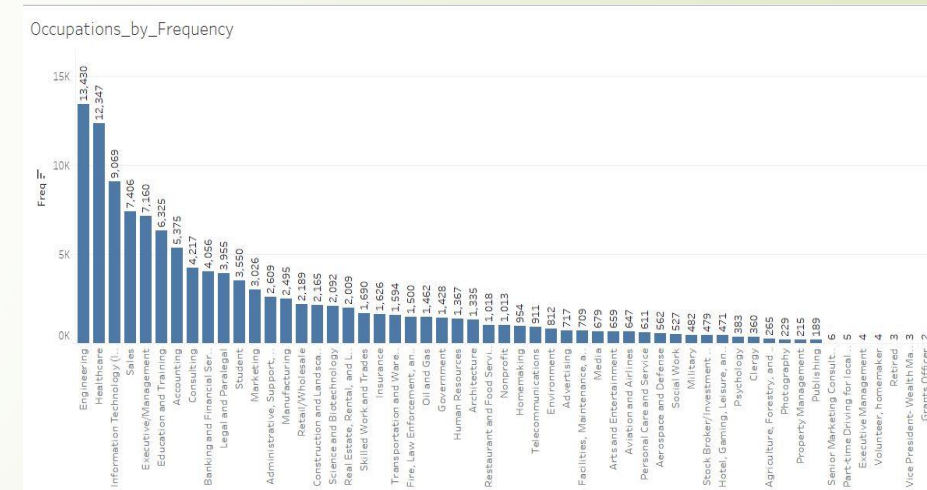
2013-2017 Bike MS Participants AND 2013-2017 Bike Teams data

Exploratory Data Analysis:

#	Parameter	Analysis Technique	Tools Used
1	Email Id	Distribution	R
2	Event Type	Frequency Distribution	R
3	Occupation	Frequency Distribution	R
4	Team Division	Binning by Clustering	R
5	Contribution by Occupation	Time Series Analysis	Tableau, R,

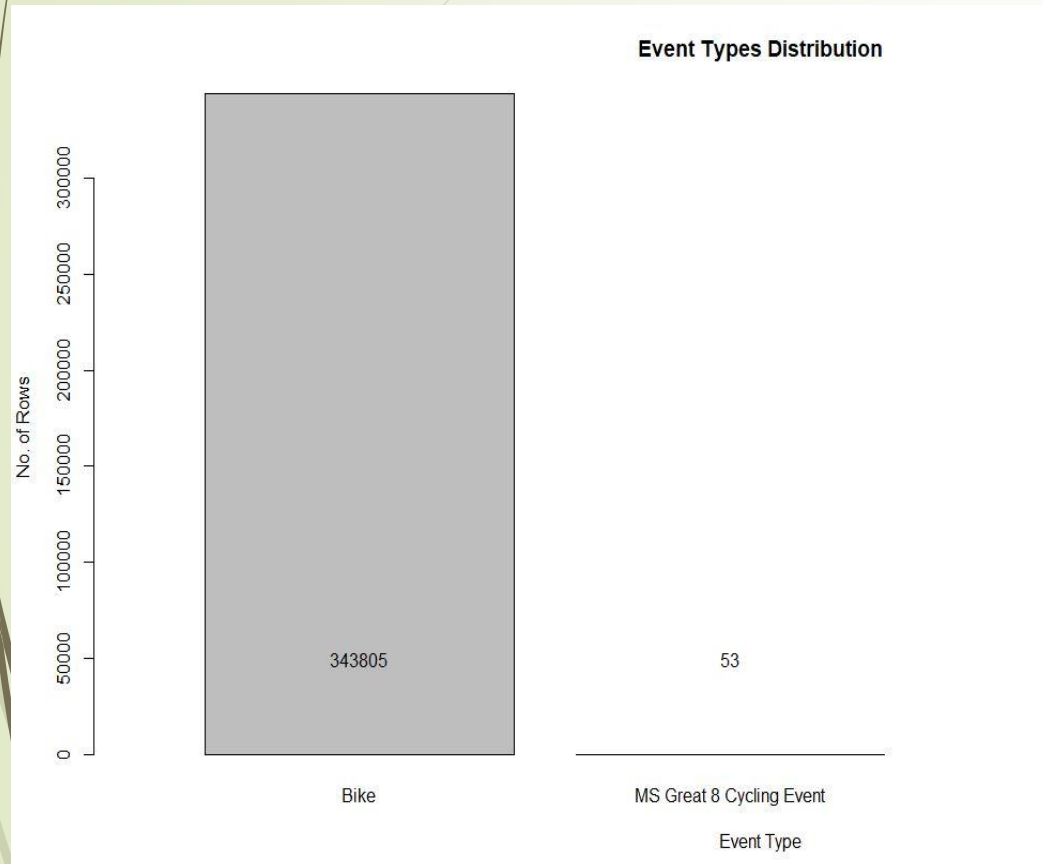
Total Revenue generated by Occupations in 5y sorted by Top 15 Occupations

	Category	x
1	Engineering	14395885.19
2	Executive/Management	12609392.65
3	Healthcare	9495687.64
4	Information Technology (IT)	7606089.71
5	Sales	7597209.04
6	Legal and Paralegal	6634729.17
7	Accounting	5695448.77
8	Consulting	5411887.07
9	Education and Training	4973761.93
10	Banking and Financial Services	4820530.99
11	Marketing	3245598.68
12	Real Estate, Rental, and Leasing	2971337.83
13	Science and Biotechnology	2184393.12
14	Manufacturing	2165160.52
15	Construction and Landscaping	2096496.36



Exploratory Data Analysis

Event Type



MS Great 8 event participation numbers are minimal when compared to Bike MS

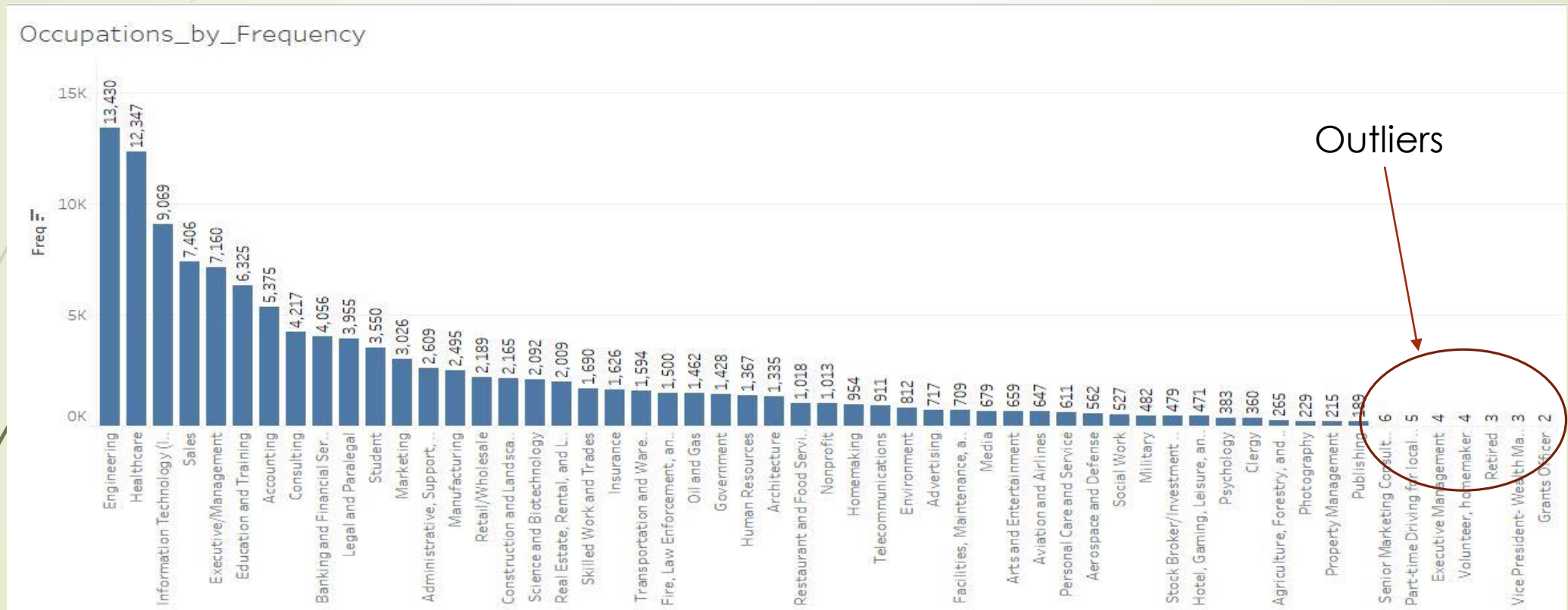
Team Division

	X	
Friends and Family	134698	Organization
Corporate	134657	Small Business
	34813	Place of worship
Family and Friends	14603	Bike Club
Corporation	12943	Ohana and Friends
Other	3548	Organization (Clubs, Civic Gr
Friend and Family	2711	Open
Organization (Clubs; Civic Groups; etc.)	2268	Open Team
Family/Friends	1320	Ohana
Place of Worship	450	Bike Shops
Beer/Brewery	411	Organization (Clubs; Civic Gr
Bike Shop	305	School
School	272	Civic Team
Club/Organization	110	Association
		Frien`s and Family
		Religious

Most of the teams came from reference through the friends and family compared to other fields.

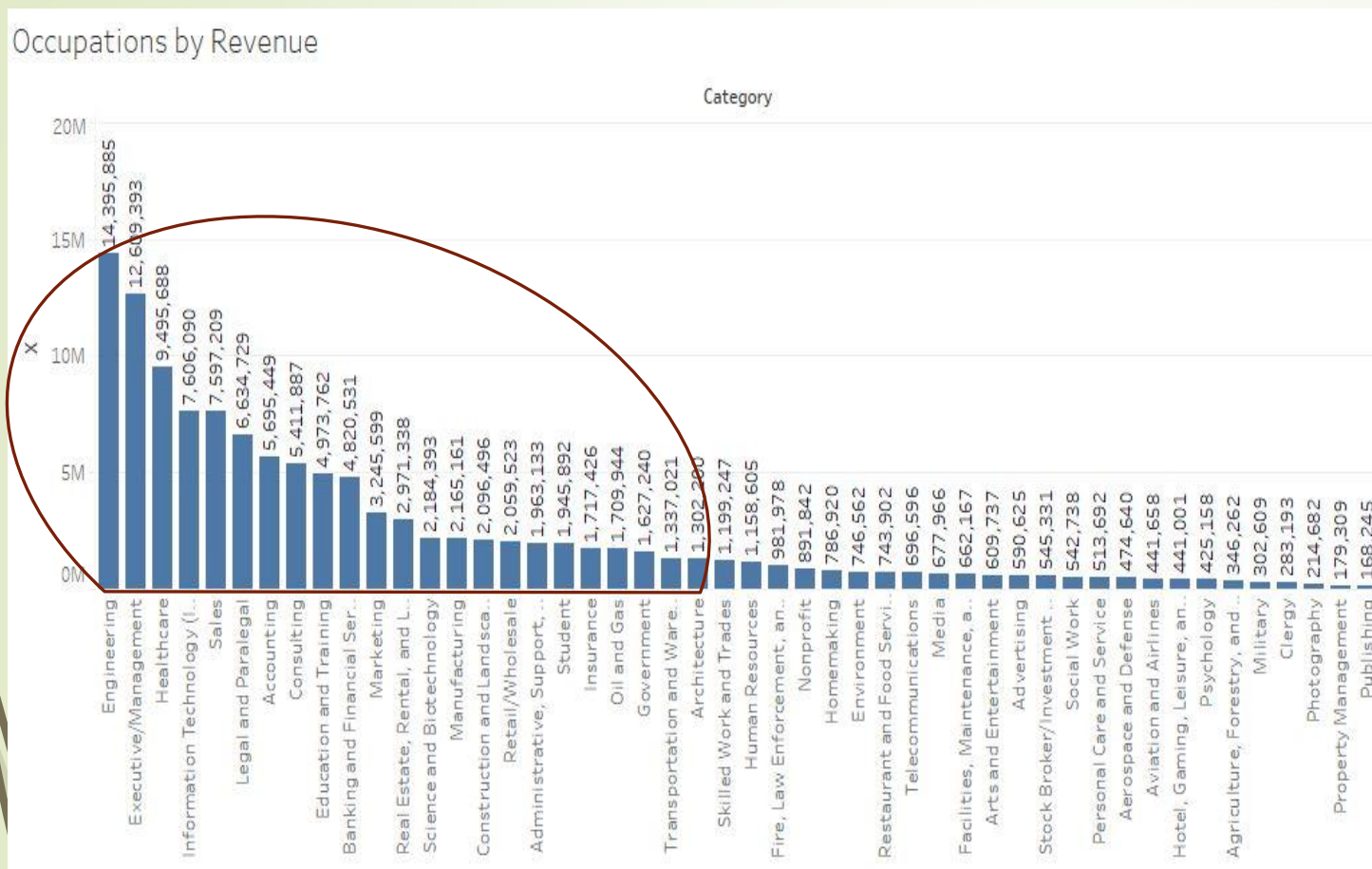
Exploratory Data Analysis

Occupation Distribution



Occupation with frequency less than 10 are omitted considering as outliers.

Exploratory Data Analysis - Occupation Distribution by revenue

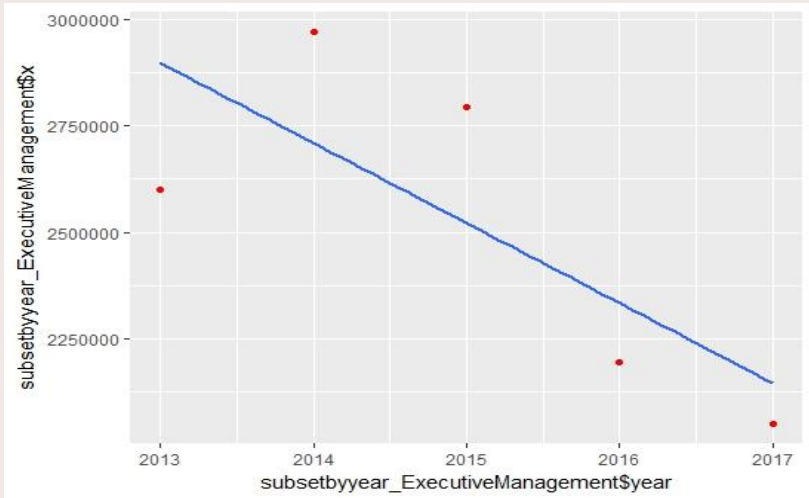
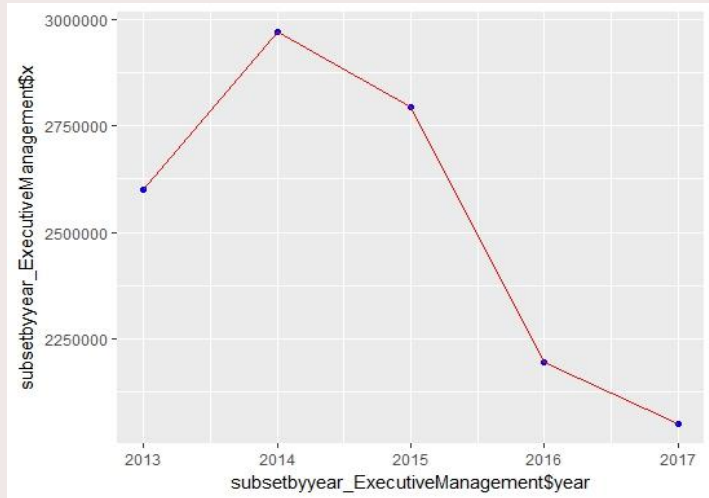


The graph shows the top performing occupations by revenue.

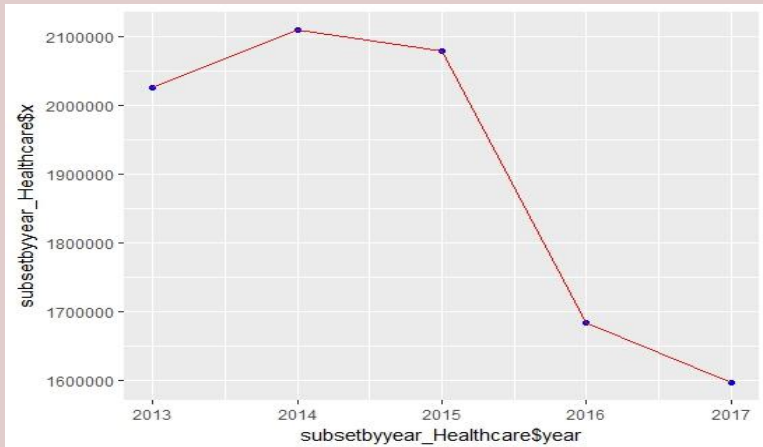
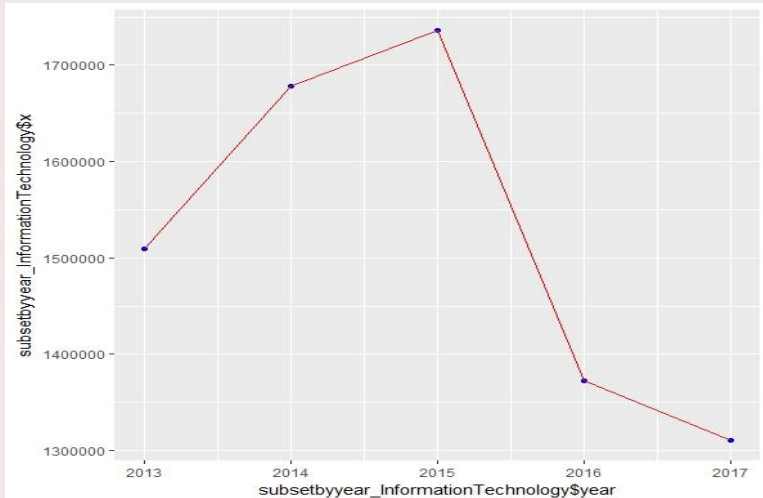
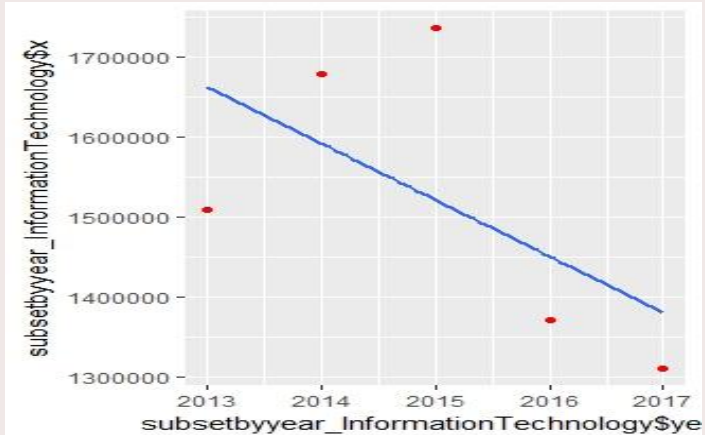
Occupation	2014 Rank	2015 Rank	2016 Rank	2017 Rank
Engineering	1	1	1	1
Executive/Management	2		2	2
Healthcare	3	3	3	3
Legal and Paralegal	4	6	6	6
Information Technology (IT)	5	4	5	5
Sales	6	5	4	4
Accounting	7	7	7	7
Consulting	8	8	8	8
Education and Training	9	9	9	9
Banking and Financial Services	10	10	10	10

Contributions by year
Sorted by top 5 occupations

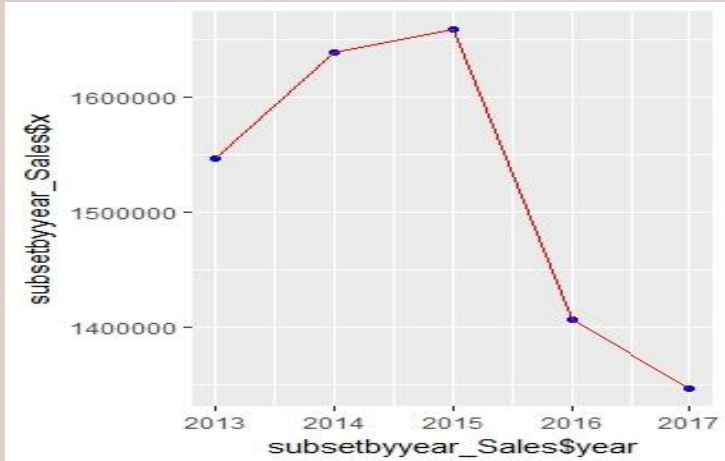
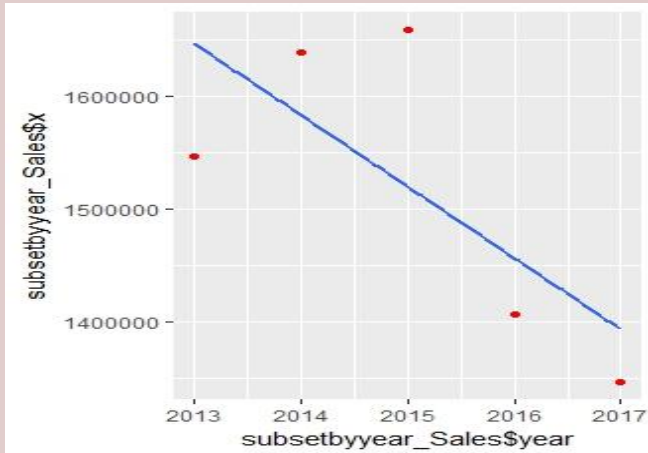
MODELLING – TIME SERIES ANALYSIS

#	Occupation	Aggregated by year map	Linear regression																								
1	Engineering	 <p>This line graph shows the aggregated salary for Engineering from 2013 to 2017. The y-axis is labeled 'subsetbyyear_Engineering\$x' and ranges from 2,500,000 to 3,250,000. The x-axis is labeled 'subsetbyyear_Engineering\$Year' and shows years 2013 to 2017. The data points are connected by a red line, showing a peak in 2014 and a sharp drop in 2016.</p> <table><thead><tr><th>Year</th><th>Salary (\$)</th></tr></thead><tbody><tr><td>2013</td><td>3,000,000</td></tr><tr><td>2014</td><td>3,400,000</td></tr><tr><td>2015</td><td>3,300,000</td></tr><tr><td>2016</td><td>2,400,000</td></tr><tr><td>2017</td><td>2,400,000</td></tr></tbody></table>	Year	Salary (\$)	2013	3,000,000	2014	3,400,000	2015	3,300,000	2016	2,400,000	2017	2,400,000	 <p>This scatter plot shows the salary for Engineering from 2013 to 2017 with a blue linear regression line. The y-axis is labeled 'subsetbyyear_Engineering\$x' and ranges from 2,500,000 to 3,250,000. The x-axis is labeled 'subsetbyyear_Engineering\$Year' and shows years 2013 to 2017. The data points are red dots, and the regression line shows a downward trend.</p> <table><thead><tr><th>Year</th><th>Salary (\$)</th></tr></thead><tbody><tr><td>2013</td><td>3,000,000</td></tr><tr><td>2014</td><td>3,400,000</td></tr><tr><td>2015</td><td>3,300,000</td></tr><tr><td>2016</td><td>2,400,000</td></tr><tr><td>2017</td><td>2,400,000</td></tr></tbody></table>	Year	Salary (\$)	2013	3,000,000	2014	3,400,000	2015	3,300,000	2016	2,400,000	2017	2,400,000
Year	Salary (\$)																										
2013	3,000,000																										
2014	3,400,000																										
2015	3,300,000																										
2016	2,400,000																										
2017	2,400,000																										
Year	Salary (\$)																										
2013	3,000,000																										
2014	3,400,000																										
2015	3,300,000																										
2016	2,400,000																										
2017	2,400,000																										
2	Executive/Management	 <p>This line graph shows the aggregated salary for Executive/Management from 2013 to 2017. The y-axis is labeled 'subsetbyyear_ExecutiveManagement\$x' and ranges from 2,250,000 to 3,000,000. The x-axis is labeled 'subsetbyyear_ExecutiveManagement\$Year' and shows years 2013 to 2017. The data points are connected by a red line, showing a peak in 2014 and a sharp drop in 2016.</p> <table><thead><tr><th>Year</th><th>Salary (\$)</th></tr></thead><tbody><tr><td>2013</td><td>2,600,000</td></tr><tr><td>2014</td><td>2,950,000</td></tr><tr><td>2015</td><td>2,800,000</td></tr><tr><td>2016</td><td>2,200,000</td></tr><tr><td>2017</td><td>2,100,000</td></tr></tbody></table>	Year	Salary (\$)	2013	2,600,000	2014	2,950,000	2015	2,800,000	2016	2,200,000	2017	2,100,000	 <p>This scatter plot shows the salary for Executive/Management from 2013 to 2017 with a blue linear regression line. The y-axis is labeled 'subsetbyyear_ExecutiveManagement\$x' and ranges from 2,250,000 to 3,000,000. The x-axis is labeled 'subsetbyyear_ExecutiveManagement\$Year' and shows years 2013 to 2017. The data points are red dots, and the regression line shows a downward trend.</p> <table><thead><tr><th>Year</th><th>Salary (\$)</th></tr></thead><tbody><tr><td>2013</td><td>2,600,000</td></tr><tr><td>2014</td><td>2,950,000</td></tr><tr><td>2015</td><td>2,800,000</td></tr><tr><td>2016</td><td>2,200,000</td></tr><tr><td>2017</td><td>2,100,000</td></tr></tbody></table>	Year	Salary (\$)	2013	2,600,000	2014	2,950,000	2015	2,800,000	2016	2,200,000	2017	2,100,000
Year	Salary (\$)																										
2013	2,600,000																										
2014	2,950,000																										
2015	2,800,000																										
2016	2,200,000																										
2017	2,100,000																										
Year	Salary (\$)																										
2013	2,600,000																										
2014	2,950,000																										
2015	2,800,000																										
2016	2,200,000																										
2017	2,100,000																										

MODELLING – TIME SERIES ANALYSIS

#	Occupation	Aggregated by year map	Linear regression																								
3	Health Care	 <table><tr><th>Year</th><th>subsetbyyear_Healthcare\$x</th></tr><tr><td>2013</td><td>2025000</td></tr><tr><td>2014</td><td>2105000</td></tr><tr><td>2015</td><td>2075000</td></tr><tr><td>2016</td><td>1685000</td></tr><tr><td>2017</td><td>1600000</td></tr></table>	Year	subsetbyyear_Healthcare\$x	2013	2025000	2014	2105000	2015	2075000	2016	1685000	2017	1600000	 <table><tr><th>Year</th><th>subsetbyyear_Healthcare\$x</th></tr><tr><td>2013</td><td>2025000</td></tr><tr><td>2014</td><td>2105000</td></tr><tr><td>2015</td><td>2075000</td></tr><tr><td>2016</td><td>1685000</td></tr><tr><td>2017</td><td>1600000</td></tr></table>	Year	subsetbyyear_Healthcare\$x	2013	2025000	2014	2105000	2015	2075000	2016	1685000	2017	1600000
Year	subsetbyyear_Healthcare\$x																										
2013	2025000																										
2014	2105000																										
2015	2075000																										
2016	1685000																										
2017	1600000																										
Year	subsetbyyear_Healthcare\$x																										
2013	2025000																										
2014	2105000																										
2015	2075000																										
2016	1685000																										
2017	1600000																										
4	Information Technology	 <table><tr><th>Year</th><th>subsetbyyear_InformationTechnology\$x</th></tr><tr><td>2013</td><td>1510000</td></tr><tr><td>2014</td><td>1675000</td></tr><tr><td>2015</td><td>1730000</td></tr><tr><td>2016</td><td>1375000</td></tr><tr><td>2017</td><td>1310000</td></tr></table>	Year	subsetbyyear_InformationTechnology\$x	2013	1510000	2014	1675000	2015	1730000	2016	1375000	2017	1310000	 <table><tr><th>Year</th><th>subsetbyyear_InformationTechnology\$x</th></tr><tr><td>2013</td><td>1510000</td></tr><tr><td>2014</td><td>1675000</td></tr><tr><td>2015</td><td>1730000</td></tr><tr><td>2016</td><td>1375000</td></tr><tr><td>2017</td><td>1310000</td></tr></table>	Year	subsetbyyear_InformationTechnology\$x	2013	1510000	2014	1675000	2015	1730000	2016	1375000	2017	1310000
Year	subsetbyyear_InformationTechnology\$x																										
2013	1510000																										
2014	1675000																										
2015	1730000																										
2016	1375000																										
2017	1310000																										
Year	subsetbyyear_InformationTechnology\$x																										
2013	1510000																										
2014	1675000																										
2015	1730000																										
2016	1375000																										
2017	1310000																										

MODELLING – TIME SERIES ANALYSIS

#	Occupation	Aggregated by year map	Linear regression																								
5	Sales	 <table><caption>Sales Data (Aggregated by year map)</caption><tr><th>Year</th><th>Sales (\$)</th></tr><tr><td>2013</td><td>1,550,000</td></tr><tr><td>2014</td><td>1,650,000</td></tr><tr><td>2015</td><td>1,670,000</td></tr><tr><td>2016</td><td>1,410,000</td></tr><tr><td>2017</td><td>1,350,000</td></tr></table>	Year	Sales (\$)	2013	1,550,000	2014	1,650,000	2015	1,670,000	2016	1,410,000	2017	1,350,000	 <table><caption>Sales Data (Linear regression)</caption><tr><th>Year</th><th>Sales (\$)</th></tr><tr><td>2013</td><td>1,550,000</td></tr><tr><td>2014</td><td>1,650,000</td></tr><tr><td>2015</td><td>1,670,000</td></tr><tr><td>2016</td><td>1,410,000</td></tr><tr><td>2017</td><td>1,350,000</td></tr></table>	Year	Sales (\$)	2013	1,550,000	2014	1,650,000	2015	1,670,000	2016	1,410,000	2017	1,350,000
Year	Sales (\$)																										
2013	1,550,000																										
2014	1,650,000																										
2015	1,670,000																										
2016	1,410,000																										
2017	1,350,000																										
Year	Sales (\$)																										
2013	1,550,000																										
2014	1,650,000																										
2015	1,670,000																										
2016	1,410,000																										
2017	1,350,000																										

PREDICTION RESULTLS

Occupation	Linear regression equation	Year	Predicted Amount
Engineering	Amount = $245701 * (\text{Year}) - 497966648$	2018	2142030
Executive/management	Amount = $187709 * (\text{Year}) - 380756779$	2018	1960017
Health care	Amount = $128556 * (\text{Year}) - 260940996$	2018	1514988
Information technology	Amount = $70309 * (\text{Year}) - 143194401$	2018	1310839
Sales	Amount = $63189 * (\text{Year}) - 128845442$	2018	1330040