

QBUS2820 - Predictive Analytics

Tutorial 2 Tasks

Task 1 - Historical Data

The sales team wants to know what the trend of average customer spending is over the past 5 years. Each year the company randomly samples 1000 customers. You have already been given one year in the "DirectMarketing.xlsx" file.

Helpfully the sales team has given you a new file called "AmountSpent_13to17.xlsx".

Use this file to calculate the average customer spend per year. Present your results as a line plot. Don't forget to label all axes, title etc. You can use the following code to label your x axis ticks:

```
plt.xticks( np.arange(1,6), ['2013', '2014', '2015', '2016', '2017'] )
```

Task 2 - Querying

Use DirectMarketing.xlsx

What is the distribution of Customer Salary?

Create a histogram of all customer's salary.

What is the distribution of Salary for High Spending Customers?

Create a histogram of customer salary for customers that spent more than 500 dollars.

Is the amount spent different between male and female customers?

1. Create a single plot with a histogram each for male and female customers
2. Calculate and report the average and variance for each customer segment