Midterm- Fall 2019

DSO 545: Statistical Computing and Data Visualization

10-10-2019

Instructions

- This is an open notes exam (1 double-sided paper). You are NOT allowed to use the Internet as a resource except for downloading the files from blackboard, and uploading your answer files back to blackboard
- Answer all questions below
- Don't change the data file names
- You have 90 mins to finish this exam
- You are NOT allowed to communicate with ANY PERSON in or outside the class during the exam period
- Suggested **max** times to spend on each case:

Case 01	Case 02
40 mins	50 mins

"I hereby certify that I have adhered to the university policies regarding ethical behavior in preparing for and completing this midterm exam. I will not discuss the exam questions and solutions with anyone in the classroom or outside the classroom via any means before Oct 10, 2019 at 2:00pm."

Name: _____

Signature : _____

Case 1: HR Analytics

In this case, we will use the "recruitment.csv" dataset. In this dataset, we have collected data about the employees' performance. The following table describes each variable in the dataset:

Variable	Description
attrition	1: employee left the company, 0: otherwise
performance_rating	employee's performance rating (1:lowest, 5:highest)
sales_quota_pct	% of sales quota
recruiting_source	the employee's recruiting channel

1. **(2 points)** What is the average attrition rate for each recruiting channel? Sort your result in descending attrition rate order.
2. **(2 points)** Create a new variable called "performance_cat". What is the average attrition rate for both employee categories (low performers vs. high performers)

Performance Rate	Performance Category
1 or 2	"Low Performance"
3, 4, or 5	"High Performance"

3. **(2 points)** Create a barchart (exact copy of the one below) to show the the average sales quota for all employees hired through different channels.

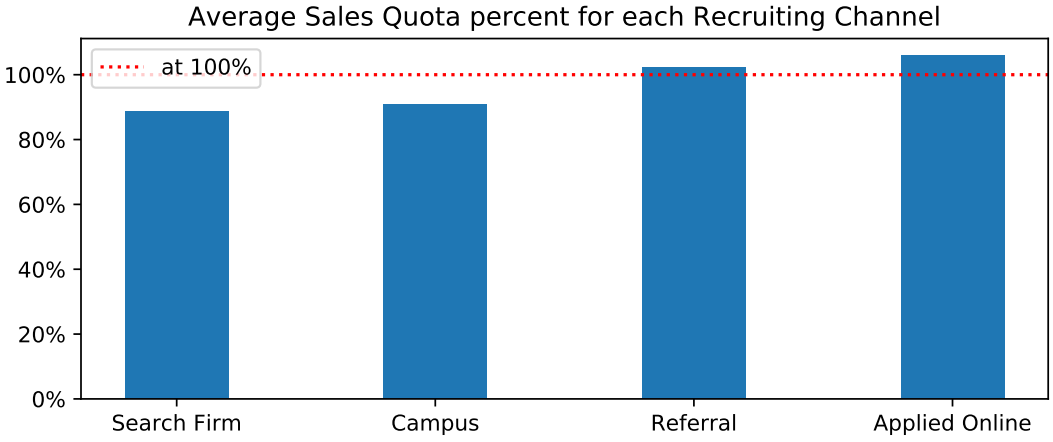
Graph Specification:

- Figure size (8,3)
- The dotted line ("red") represents the 100% line
- Use default colors for the bars
- The width of the bars is 0.4

```
## <BarContainer object of 4 artists>
```

```
## [<matplotlib.axis.YTick object at 0x12817f5d0>, <matplotlib.axis.YTick object at 0x128175d50>, <matpl
```

```
## [Text(0, 0, '0%'), Text(0, 0, '20%'), Text(0, 0, '40%'), Text(0, 0, '60%'), Text(0, 0, '80%'), Text(
```



Case 2: Financial Savings Bank

Suppose that you work at a bank called "Financial Savings Bank", and you would like to compare its performance against it's peers (Bank1, Bank2, . . . , Bank9). You have data on the bank index (satisfaction score) over time for your bank as well as your competitors given in the dataset "bankindex.csv".

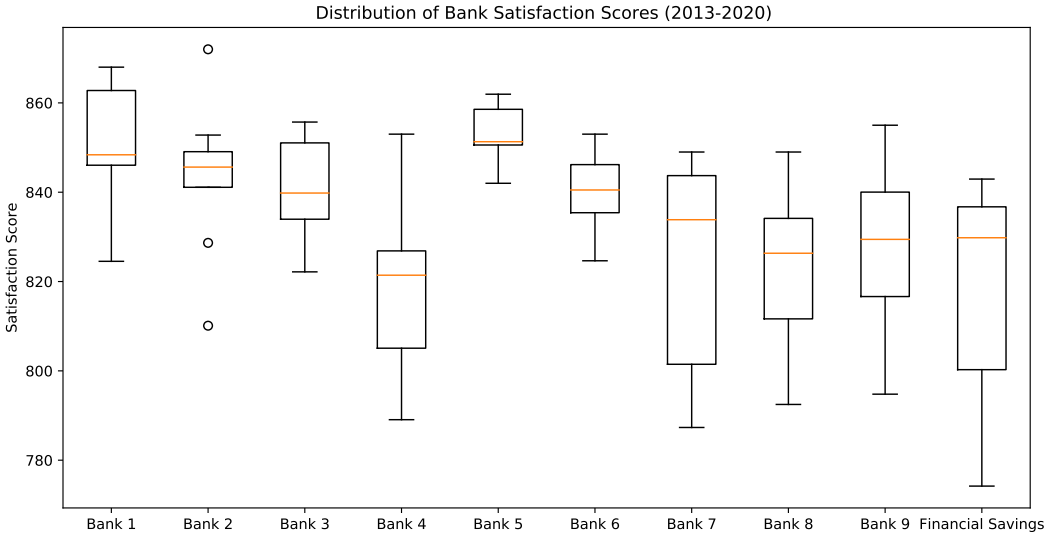
Variable	Description
Bank	bank name
Year	year
Score	satisfaction score

4. **(2 points)** Create a side-by-side boxplot (see below) to show the distribution of satisfaction scores for your bank ("Financial Savings") and your competitors.

Graph Specification:

- Figure size is (12, 6)
- Use default colors

```
## {'whiskers': [<matplotlib.lines.Line2D object at 0x1282ce990>, <matplotlib.lines.Line2D object at 0x
```



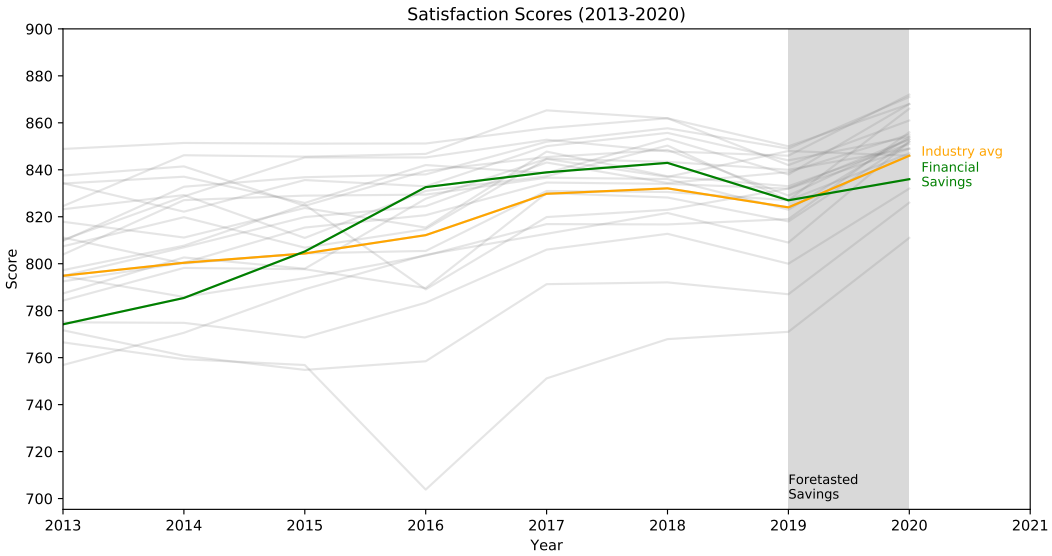
5. **(2 points)** Reproduce an exact copy of the following plot that compares your bank's performance to your competitors as well as the industry average pver the years (2013-2020) (Note that the industry average is added at the very end of the dataset).

Graph Specification:

- Figure size (12,6)
- All competitor banks are shown using "grey" lines with alpha = 0.2
- Financial Savings bank is "green"
- Industry average is "orange"
- The font size for the xticks labels is 9 or whatever is good for your graph

```
## (2013, 2021)
```

```
## [<matplotlib.axis.YTick object at 0x128b08dd0>, <matplotlib.axis.YTick object at 0x128b08510>, <matpl
```


Satisfaction Scores (2013-2020)
900
880
860
840
820
800
780
760
740
720
700
Score
2013
2014
2015
2016
2017
2018
2019
2020
2021
Year
Industry avg
Financial
Savings
Foretasted
Savings