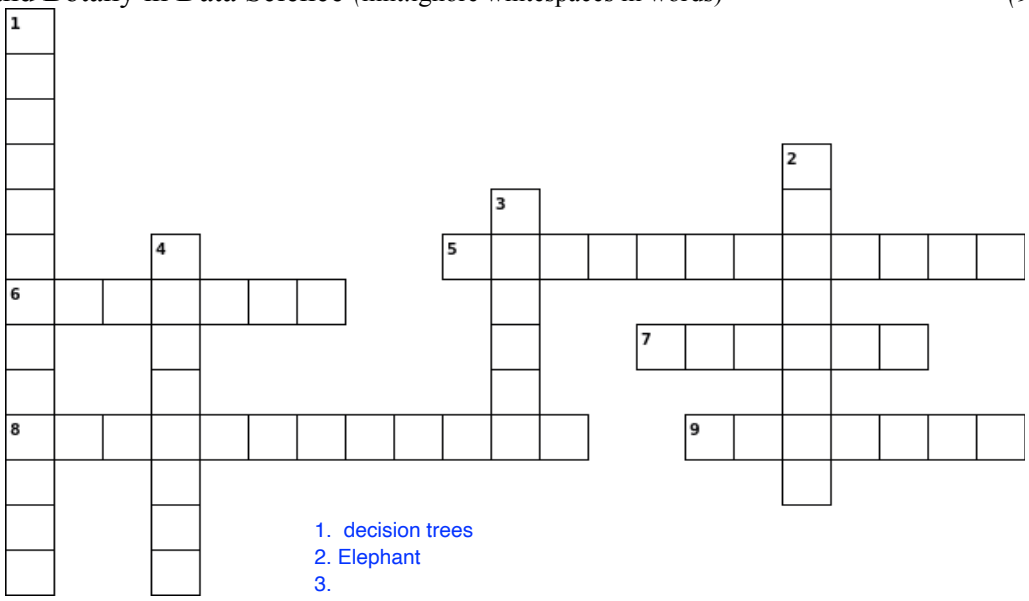
Data Science Quiz

Question 1

Zoology and Botany in Data Science *(*hint:ignore whitespaces in words) *(9 points)*



1. decision trees
2. Elephant
3.
4. Mobydock
5. Larrythebird
6.
7. Python
8. RandomForest
9.

Down

1. Are prone to overfitting

2. Has a memory like Postgres

3. Have six fingers

4. Friendly marine mammal that carries containers

Across

5. Microblogging mountain bluebird named after basketball legend

6. Feline with 8 branches

7. Nonvenomous ambush predator

8. If you spill a basket of acorns, you get this

9. Data management the natural way

Question 2

for line in open…. :
columns= lines…..
total += columns[2]

print('result:%s births total' %(year))

Find 10 bugs: *(10 points)*

```
for year in range(1890, 2015, 1):
    total = 0
    filename = 'names\yob{year}.txt'
    for line in open(filename, 'w')
        columns == line.split(',').strip()
total =+ columns[2]

print("Result: {} births total".format(year))
```

Question 3

What do the following bash commands do? *(6 points)*

`ls -a`	1. show all the files including the hidden ones. 2. as a super admin to delete rf 3. Chmod 700 (chmod a+rwx,g-rwx,o-rwx) sets permissions. * means everthing 4.
`sudo rm -rf /`	
`chmod 700 *`	
`grep print *.py | wc -l`	

Question 4

Write an SQL query that extracts the 10 most frequently occurring items in the *'subject'* column from the table *'data_scientists'*, but only consider students with the column *'python'* being 1 or higher. Output results in descending order. *(10 points)*

```
SELECT subject, COUNT(subject) as count FROM
data_scientist
WHERE python >=1
GROUP BY subject
ORDER BY count DESC
LIMIT 10;
```

Question 5

Name the functions. *(9 points)*

$P(A\|B)=\frac{P(B\|A)P(A)}{P(B)}$	
$\frac{1}{N}\sum_i (y_i - y_i^{true})^2 + \lambda \sum_j b_j^2$	1. Naiver Bayers. 2. L2 regulation 3. Logistic Regression/ Sigmoid
$\frac{1}{1+e^{-x}}$	

Question 6

Write 2 items you could import from each Python module. *(10 points)*

`pandas`	`DataFrame`
`random`	1. pandas DateOffset
`numpy`	2. random = choice, choices 3. numpy. add_newdoc, add_docstring 4. seaborn.heatmap, barplot
`seaborn`	5. os.listdir
`os`	

Question 7

Which strings does the Regular Expression `'R[oau]\w+e'` match? *(8 points)*

Rome	rose	Rue	Dome
Rhizome	Rhizome	Ru\w+e	Raave

Raave
Rome

Question 8

Write the Docker one-liners to do the following *(10 points)*

Run and start a standard python container with the name `my_python`	
Interact with `my_python`	
Calculate 4+4 in `my_python`	
Stop `my_python`	
Delete `my_python`	

Question 9

Match pairs. *(7 points)*

postgreSQL	MongoDB
`Table`	`use sales`
`Row`	`show collections`
`\l`	`Document`
`\dt`	`db.sales.find`
`\c sales`	`Collection`
`SELECT * FROM sales;`	`db.sales.distinct('client')`
`SELECT DISTINCT client FROM sales;`	`show dbs`

1. Table = Collection
2.Row = Document
3. l = show dbs
4. dt = show collections
5. c sales = use sales
6. Select ... = db.sales.find
7. Select Distinckt ... = db.sales.distinct()