Python

1. Explain how the *lambda* function works

   Ans: anon function to make python functions more concise since don't need to define it.

   ```
   Eg: def identity (x):
           return x
   Whereas: lambda x, y: x + y + 1
   ```

   (where lambda ___ < arguments)

2. What happens when you multiply the following by 3
   a. List: [2, 3, 2]*3 = [2, 3, 2, 2, 3, 2, 2, 3, 2]
   b. String: '342'*3 = 342342342
   c. Dict: cannot be done?

3. What is the difference between lists and arrays?

   Ans: they're two different types in python (both are used to store data, both are mutable, both are indexed, both can be iterated through, both can be sliced).

   - Lists can contain elements of different data types, python *array* module's "array" can only contain elements of same data type (numpy array can contain different data types if you set dtype = 'object' for that array)
   - Arrays can store data more compactly / efficiently
   - Differ in terms of operations you can perform on them.
     - Arrays are more efficient with numerical operations. *Note: numpy arrays are just one type of arrays meant for robust numerical computations cuz builtin python lists don't support this protocol (needs extra step to loop over each item in the list then save to another list)*

       ```
       Eg. cost1 = array ([4,8,12]), cost2 = [4,8,12]
       divided_cost = cost1/2 > print > [2,4,6]
       divided_cost2 = cost2/2 > print > type error
       ```

   - Arrays need to be declared, lists don't

4. How do you slice a list?

   ```
   Ans: nums = [10, 20, 30, 40, 50, 60, 70, 80, 90]
   some_nums = nums[2:7]
   ➢ [30, 40, 50, 60, 70]
   ```

5. How would you remove duplicate elements from a list?

   my_list = [1, 1, 2, 3, 2, 2, 4, 5, 6, 2, 1]

   a. Using set method:

      ```
      my_final_list = set(my_list)
      ```

   b. Using temporary list:

      ```
      temp_list = []
      for i in my_list:
          if i not in temp_list:
              temp_list.append(i)
      ```

   c. Using numpy unique() method:

      ```
      Import numpy as np
      my_final_list = np.unique(my_list).tolist()
      ```

   d. Using pandas methods

```
Import pandas as pd
my_final_list = pd.unique(my_list).tolist()
```

e. Using enumerate() and list comprehension

```
My_final_list = [i for j, i in enumerate(my_list) if i not in my_list[:j]]
```

6. How would you sort a dictionary by key, alphabetically?

```
orders = {
        ‘cappucino’: 54,
        ‘latte’: 56,
        ‘espresso’: 72,
        ‘americano’: 48,
        ‘cortado’: 41}
```

Sort by key alphabetically:

```
sort_orders_value = sorted(orders.items(), key = lambda x: x[0], reverse=False)
for i in sort_orders_value:
        print(i[0], i[1])
```

- Americano 48
- Cappuccino 54
- Cortado 41
- Espresso 72
- Latte 56

Lambda x: x[0]

*sorting mechanism that identifies element to be sorted, in this case 1st element (indexing) in each dictionary element eg ‘cappucino’: 54, 1st element is ‘cappucino’

Sort by value:

```
sort_orders_value = sorted(orders.items(), key = lambda x: x[1], reverse=True)
for i in sort_orders_value:
        print(i[0], i[1])
```

- Espresso 72
- Latte 56
- Cappuccino 54
- Americano 48
- Cortado 41

7. Check if a string only contains

a. Numbers

```
Initial_string = ‘_____’
If initial_string.isdigit():
        Print(“string contains only numbers”)
Else:
        Print(“string doesn’t only contain numbers”)
```

_.isdigit() returns True if all characters in string are digits

b. Letters

```
Initial_string = ‘_____’
Intial_string_lowercase = intial_string.lower()
#converts all letters to lowercase

Contains_letters = initial_string_lowercase.islower()
```

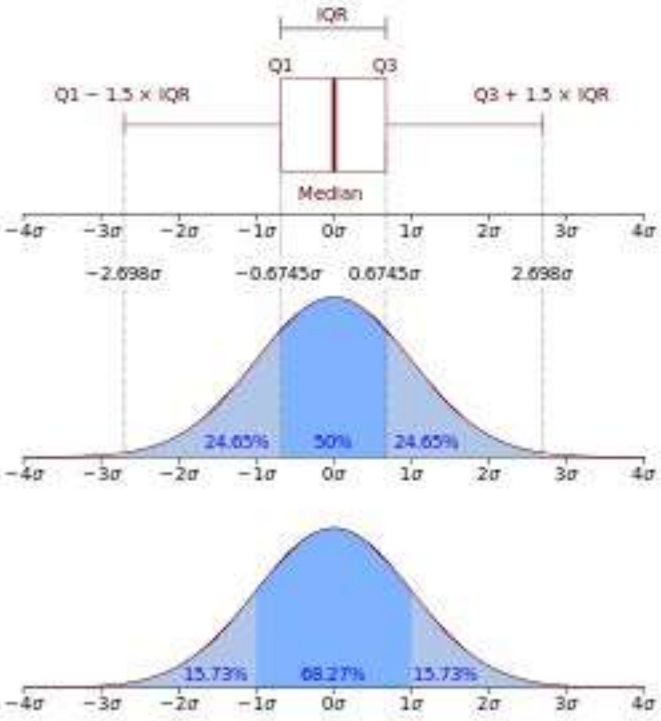
Can do this with __.upper() and __.isupper() also

c. Numbers and Letters

same as above but use __.isalnum() which will return True if string only contains alphanumeric characters (numbers/alphabets)

Statistics

8. In a box-and-whiskers plot, how do you define
   a. The edges of the box = 25th and 75th percentile of population
   b. The length of the whiskers = 1.5*IQR (https://towardsdatascience.com/why-1-5-in-iqr-method-of-outlier-detection-5d07fdc82097)



9. In layman terms, what do the following represent?
   a. Variance
   b. Standard deviation = reflects variability in distribution (this is squared value of (a), sd is in same units as original population data values)
   c. Median = 50th percentile of population (less prone to skewing compared to mean value)
10. What is the difference between a probability density function (pdf) and cumulative distribution function (cdf)? Use the normal distribution for your answer, to keep things straightforward.

    Ans: pdf area under curve is 1, not the case for cdf; pdf y-value is prob for that specific corresponding x value, but for cdf the y-value is accummulated prob for all x-values up to that value.
11. What is the purpose of calculating a z-score for conducting statistical tests?

    Ans: z-score measures how many sd's a data pt is away from the mean, if z-score is zero, it is the mean. Helpful to compare data from diff populations (with diff means, diff sd's)

12. How does a least squares regression model arrive at the best fit line?

    Ans: it calculates, using all the data points present, which line will fall in a position such that the squared distance (residuals) of each point to the line is minimized (predicted value vs actual value – diff is minimized)

13. What does the height of the bars represent, in a binomial distribution pdf?

    Ans: probability (y value) of getting a certain result (corr x value)

14. In what cases should you use a linear regression model?

    Ans: when there’s a linear relationship between 2 variables