**Tarea 1.**

En un documento de Google Colab(.ipynb) desarrollar en python(3.7 o superior) lo siguiente:

1. **(1pt)**Desarrollar el código para cargar archivos en Google Colab.
   1. obtenga los datos desde python del archivo Vibrio\_cholerae.txt en modo lectura.
   2. Guarde la cadena dentro del archivo en una variable.
2. **(1pt)**Desarrolle una función de python para obtener la cantidad de nucleótidos que tiene una cadena en total, y obtenga el porcentaje de adeninas, guaninas, citosinas y timinas que contiene la cadena.
   1. **Nombre de la función:** porcentaje\_de\_nucleotidos.
   2. **Parámetros a ingresar:** cadena de nucleótidos y nucleótido a buscar el porcentaje.
   3. **Retornar:** porcentaje de nucleótido calculado.
3. **(1pt)**Desarrolle una función de python para obtener solamente 100 nucleótidos de una cadena, en la cuál, se agregue el índice de inicio y automáticamente debe obtener los 100 nucleótidos a partir de allí. Obtenga la cadena complementaria correspondiente, considerando que la lectura de una cadena es de 5’ a 3’.
   1. **Nombre de la función:** buscar\_complemento.
   2. **Parámetros a ingresar:** cadena de nucleótidos e índice de inicio.
   3. **Retornar:** diccionario con cadena de muestra de 100 nucleótidos y su complemento.
   4. **Formato de diccionario:** retorno = {‘cadena’:’’, ‘complemento’:’’}

Por ejemplo, para una cadena más corta de 48 nucleótidos:

5’ **AGCTGTGCGAGCGAGTGTGCAGAGCGTGAGCGTCGAGTGAGCGCGTGA** 3’

su complemento es:

5’ **TCACGCGCTCACTCGACGCTCACGCTCTGCACACTCGCTCGCACAGCT** 3’

1. **(1pt)**Desarrolle una función de python que encuentre los índices de una cadena en los cuales inicia una subcadena(o palabra) almacenada en una variable.
   1. **Nombre de la función:** buscar\_indices
   2. **Parámetros a ingresar:** cadena de nucleótidos, y palabra a buscar dentro de la cadena.
   3. **Retornar**: arreglo de índices dónde se encuentra el inicio de cada palabra.

Por ejemplo:

subcadena = “AGCT”

5’ AGCTGCGAGCTACGTGAGCTCGAAATAAAGCTAGCTCGATA 3’

**índices en donde aparece: [0, 7, 16, 28, 32]**

1. **(1pt)**Desarrolle una función de python que encuentre el total de subcadenas(o palabras) que aparecen en una cadena.
   1. **Nombre de la función:** total\_de\_palabras
   2. **Parámetros a ingresar:** cadena de nucleótidos, y palabra a buscar dentro de la cadena.
   3. **Retornar:** número entero del total de veces que aparece la palabra en la cadena.

subcadena = “AGCT”

5’ AGCTGCGAGCTACGTGAGCTCGAAATAAAGCTAGCTCGATA 3’

total: 5

1. **(2.5pt)**Desarrolle una función en python que genere una cantidad de subcadenas de *n* nucleótidos a partir de una cadena principal. Obtenga el porcentaje de nucleótidos que se encuentran en cada una de las subcadenas generadas. Almacene los porcentajes de manera ordenada en una lista.
   1. **Nombre de la función:** split\_equal\_strings*.*
   2. **Parámetros a ingresar:** cadena de nucleótidos, cantidad de subcadenas a generar.
   3. **Retornar:** lista de porcentajes de nucleótidos encontrados.
2. **(2,5pt)**Busque la cantidad de veces que se repiten palabras de *n* nucleótidos dentro de la cadena Vibrio\_cholerae.txt.
   1. **Nombre de la función:** search\_words
   2. **Parámetros a ingresar:** cadena de nucleótidos, tamaño de palabras a buscar.
   3. **Retornar:** Diccionario ordenado con las palabras encontradas, asociadas a la cantidad de veces que se encontraron las palabras.

**A entregar: Tarea Individual**

1. Trabajar en Google Drive(Google Colab) su tarea.
   1. Describa cada problema en un bloque de código diferente de manera ordenada.
   2. Documentar textualmente lo que hace cada una de sus funciones.
2. Al finalizar ir a File/Download/download **.ipynb**
3. guarde el documento con su nombre completo, por ejemplo. Jorge\_Augusto\_Balsells\_Orellana.ipynb, y asegúrese que funcione bien, cargando de nuevo el documento a google drive y ejecutarlo nuevamente antes de entregarlo.
4. Ya con completa seguridad que funciona, enviar el archivo al correo [jbalsells@digi.usac.edu.gt](mailto:jbalsells@digi.usac.edu.gt) antes del 25 de Marzo de 2022 a las 11:59:59 PM.

**Nota de la tarea: 10 puntos.**

**en caso de detectar copias, automáticamente ambas tareas serán anuladas.**