

```

import pandas as pd
import numpy as np
from sympy import *
from sympy.utilities.lambdify import lambdify
from scipy import integrate
import matplotlib.pyplot as plt

movie_df = pd.read_csv("movie_coefficients_ratings.csv")

movie_df.head()

x = symbols("x")

def get_function(original_coefficients):
    x_vec = np.array([1, x, x**2, x**3, x**4, x**5, x**6])
    return np.dot(x_vec, original_coefficients)

def get_derivative_coefficients(original_coefficients):
    f = get_function(original_coefficients)
    f_prime = diff(f)
    coeffs = Poly(f_prime, x).coeffs()
    return coeffs

def get_degree(derivative_coefficients, numMonths):
    roots = np.roots(derivative_coefficients)
    roots = [root for root in roots if root.real == root and 10 < root < numMonths]
    return len(roots) + 1

def get_average_arc_length(original_coefficients, numMonths):
    f = get_function(original_coefficients)
    f_func = lambdify(x, f)
    f_prime = diff(f)
    f_prime_func = lambdify(x, f_prime)
    s, error = integrate.quad(lambda x: np.sqrt(1 + f_prime_func(x)**2), 1, numMonths)
    a = numMonths
    b = f_func(numMonths) - f_func(1)
    c = np.sqrt(a**2 + b**2)
    return s / c, b

toy_story_coefs = np.array([ 4.21326227e+00, -2.41242972e-02,  8.52165923e-04, -1.27536182e-06,
                             8.75568436e-08, -2.77284980e-10,  3.29410018e-13])
jumanji_coefs = np.array([ 3.68677941e+00, -1.83283020e-02,  2.69105349e-04, -3.38954003e-06,
                             2.57311796e-08, -9.12554907e-11,  1.18905136e-13])
mulan_coefs = np.array([ 3.83629105e+00, -4.03490174e-06, -7.17732545e-05, -7.31736136e-07,
                             1.98457139e-08, -1.08066369e-10,  1.80709578e-13])

```

```

-----, -----, -----,
iron_man_coefs = np.array([ 4.06817623e+00, -1.91306460e-02,  3.05597485e-04,  9.87567393e-06,
                           -2.98324315e-07,  2.57940074e-09, -7.28640065e-12])
skyfall_coefs = np.array([ 3.79173621e+00,  7.17252354e-03, -2.07269423e-04, -1.70677785e-05,
                           7.71114236e-07, -1.12561103e-08,  5.68371016e-11])
casino_royale_coefs = np.array([ 2.87488363e+00, -9.75469962e-03,  2.29905091e-05,  8.0514316e-07,
                                -1.26852945e-07,  6.97615789e-10, -1.31230345e-12])
ice_age_coefs = np.array([ 3.60849605e+00,  3.06628305e-03, -8.34394842e-05, -1.69765647e-07,
                           1.78168745e-08, -1.41921465e-10,  3.26086980e-13])

```

```

def doMovieOld(coefs, numMonths=250):
    derivative_coefs = get_derivative_coefficients(coefs)
    degree = get_degree(derivative_coefs, numMonths)
    arc_length = get_average_arc_length(coefs, numMonths)
    return degree, arc_length

```

```
get_average_arc_length(casino_royale_coefs, 180)
```

```

def get_movie_variables(months):
    movie_ids = []
    arcs = []
    bs = []
    for i in range(movie_df.shape[0]):
        row = movie_df.iloc[i]
        coefs = row[2:-1]
        #numMonths = row[1]
        movie_id = row[0]
        #arc, b = get_average_arc_length(coefs, numMonths)
        arc, b = get_average_arc_length(coefs, months * 1.0)
        movie_ids.append(movie_id)
        arcs.append(arc)
        bs.append(b)
    ids_pd = pd.DataFrame(movie_ids)
    arcs_pd = pd.DataFrame(arcs)
    bs_pd = pd.DataFrame(bs)
    all = pd.concat([ids_pd, arcs_pd, bs_pd], axis=1)
    all.columns = ["movie_id", "s/c", "d_height"]
    all.to_csv("movie_variables_" + str(months) + "_months.csv", index=False)

```

```

for i in [1, 4, 12, 24, 60]:
    get_movie_variables(i)
    print(i)

```

```
movie_variables_df.to_csv("movie_variables_12_months.csv", index=False)
```

```
movie_variables_df
```

```
movies_w_genres = pd.read_csv("movies_w_genres_after85.csv", index_col=0)
```

```
movies_w_genres = pd.read_csv('movies_w_genre_after96.csv', index_col=0)
movie_variables_df = pd.read_csv("movie_variables.csv", index_col=0)
movie_variables_after_1996_df = movie_variables_df[movie_variables_df.index.isin(movies_w_genres)]
movie_variables_after_1996_df.to_csv("movie_variables_after_1996.csv")
```

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.