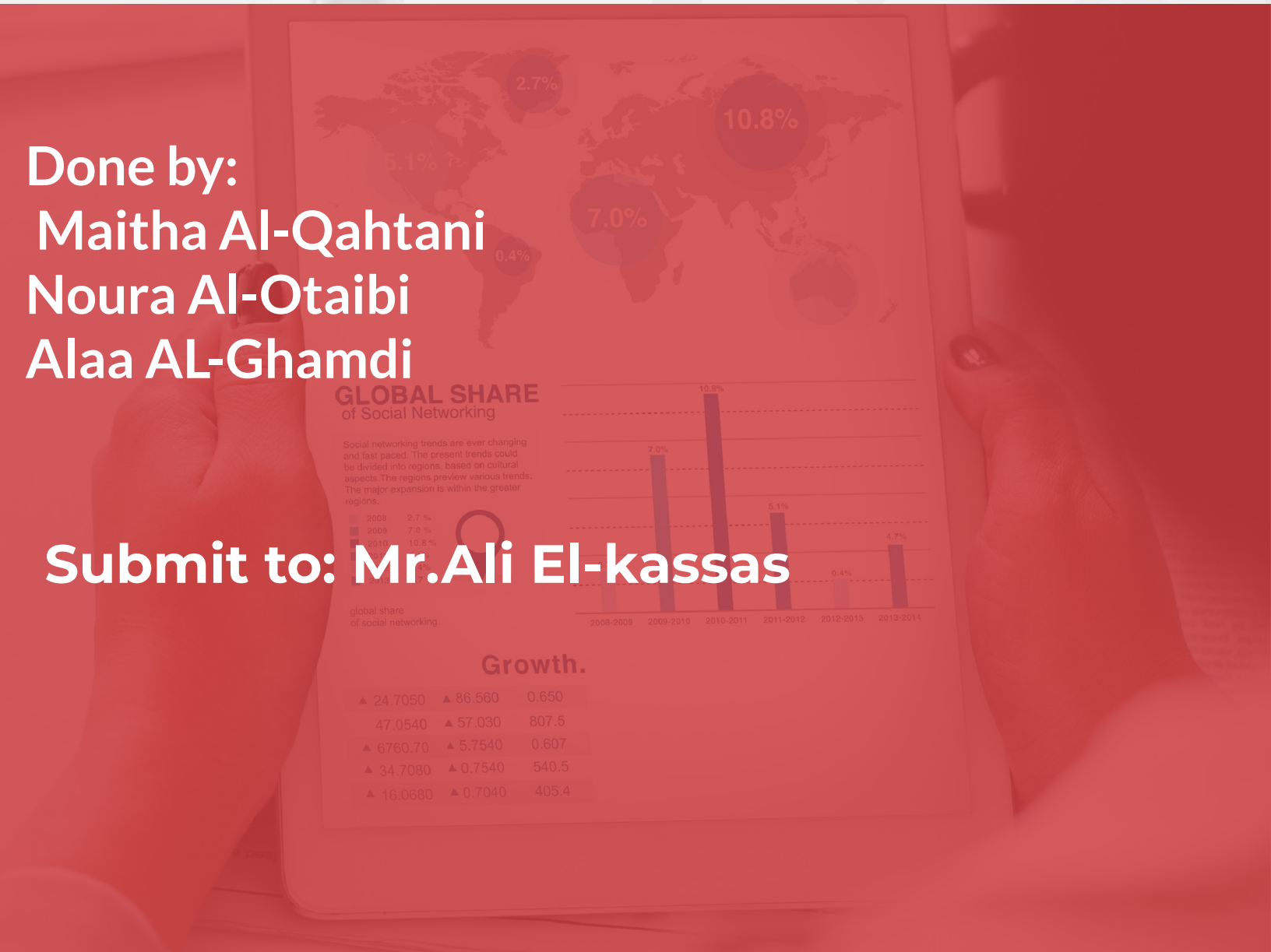
How popular will this video be on YouTube?

FINAL PROJECT



Done by:
Maitha Al-Qahtani
Noura Al-Otaibi
Alaa AL-Ghamdi

Submit to: Mr.Ali El-kassas

Introduction :

YouTube is the world's second largest search engine. As one of the most popular Art & Entertainment content-providing systems, YouTube spans across cultures and nations, with its trending videos viewed by a wide range of audiences around the globe. Thus, learning about characteristics of trending YouTube videos .

Data Description:

The dataset that we will use is obtained from Kaggle here. It contains data about trending videos for many countries. Here we will analyze USA and Canda trending videos.

Initial Column Observations:
video_id: contains alphanumeric code for video identification. However, this is not helpful for our data exploration/analysis
trending_date: contains the date the video started trending in YYDDMM format
title: contains the title of the video.
channel_title: contains the title of the channel
category_id: contains the id number for each category. We will need to match the id number to its corresponding category name
publish_time: contains the date and time the video was published. It is formatted differently from trending date, which we plan to clean
tags: contains the tags in one long string, which we will need to separate into a list
views, likes, dislikes, comment_count: contains numerical values
thumbnail_link: contains url for picture of the thumbnail.
comments_disabled, ratings_disabled,
video_error_or_removed: contains boolean values
description: contains the video description as a string. Some contain non-ascii characters, emojis, and urls which we will need to remove if we use

Algorithms:
- Import necessary libraries and read data set.
- Convert some rows into more than one type.
- Added a new feature.
- Data Analysis and Visualization.
- Data cleaning.
- Check null values.
- Check duplicated values.
- Describtion of numerical columns.
- Evaluation metrics is: Recall.
- Use more than one modal to choose the best degree.

Tools:
1.Numpy
2.Pandas
3.Matplotlib
4.Seaborn
5.Sklearn
6.collections
7.Statsmodels.
8.WordCloud
9.random

Communication:
Presentation and Visualization .