GENERELL: Mitschreiben was wir gemacht haben!!!

Wenn Datensatz fehlerhaft war Warum?

Wenn Methode nicht funktioniert hat warum?

Wie hat man es probiert?:..

Und Ideen die wir währenddessen haben immer mitschreiben. Was kommt noch? Was muss noch in das Dokument?

 -------------------------------------------------------------------------------------------------------------------------------------

 Bis zum 13.02

Wie erkennt man Künstler in dem Prompt/ String?

1. Duplikate erkennen und bei kleinen Datensatz auf 1.5M großen 1.8M unique prompts kommen? Wie gruppieren-> Doku? MORITZ  
     
   File
2. Welchen Referenzdatensatz nehmen wir? MORITZ

Kann auch wahrscheinlich andere sein als die Excel, hauptsache vollständig.  
Geht erstmal darum "alle" zu erkennen und noch nicht das Mapping zu Epoche, Zeit. Wenn beides auf einmal umso besser

Vollständige epochenliste mit zeit

https://en.m.wikipedia.org/wiki/Periods\_in\_Western\_art\_history

Vollständige stilliste

<https://en.m.wikipedia.org/wiki/List_of_art_movements>

https://magazine.artland.com/art-movements-and-styles/

Am besten zusammen aber einzeln auch ok erstmal der Vollständigkeit halber

<https://weirdwonderfulai.art/resources/disco-diffusion-70-plus-artist-studies/>

<http://www.artcyclopedia.com/artists/AZ.html>

1. Wie kann man die prompts aus dem parquet file nehmen und die Künstler in dem string erkennen -> code  
   Prüfen, dass nur Nachname geht, Vor und Nachname, Groß und Kleinschreibung, leichte Verschreiber wie Vinncent van Gogh muss auch gehen in der Kombination GEORG

Named Entity Recognition?

Part of speech tagging

--------------------------------------------------------------------------------------------------------------------------------------

SPÄTER

4. In wie vielen Prompts kommen 2 oder mehr Künstler vor ? ---> Georg  
 5. In wie viel Prozent der Prompts tauchen mindestens ein Künstler auf?

Wer kommt zusammen mit anderen vor Kookurrenz ---> Georg

1. Schauen wir wie Nutzer user\_id innerhalb von kurzer Zeit timestamp iterativ probiert haben bessere Ergebnisse bekommen oder wird das schon durc 1. Rausgefiltert mit unique?

Auf jedengall mal schauen Zusammenhänge zwischen einzelnen Rows group by’s machen?

Was ist dann das Endergebnis was zufriedenestellen dist Welcher Künstelr, stil, epoche?

1. Quellenkritik Wie viele unique User? Welcher Zeitraum? Ist der klein? Nur Stable-Didffusionserver als Quelle
2. Seiten drüber schrieben was ist parquet? Was macht das gut? Was schlecht? Wie ist support/ community? Performance=

Generell in Textabstract z Bsp: Was machen wir in DH? Mit maschinellen Verfahren Daten annotieren, um dann auf den Daten statistische Auswertung zu machen und damit dann die Hypothese signifikant zu validieren oder auch nicht.

Struktur:

Einleitung

Verwandte Arbeiten

Methodik

Experimentdurchführung

Diskussion

Explorativ ist gut um Daten zu versteheen aber

Aber an sich für uns besser thesen zu testen und deduktiv arbeiten

--

Extraktion unique prompts und der künstlerliste

Fuzzy search hat zu lange gedauert für alle Künstler(erstmal, ohne Parallelisierung)

Exact matching. Gut

Stellt sich raus die famous typen. primör Internetkünstler

Mapping mit stilen(epochen) nehmen.

Coocurenz mit vershciedenen stilen(epochen)

Wie sind die häufigkeiten innerhalb der stilgruppe vertreten? Ein künslter macht 70% der expresionisten aus

* 1. Einleitung Was Ziel … -> Expose
     1. Was gibt es für andere Paper
  2. Welche Daten?
     1. Quellenkritik PRIO Excel, aber auch Poloclub  
        Vorteile Datensätze etc. MORITZ
     2. Mapping Künstler Stil ist neuer Datensatz. Aus Künstler mit Tag muss man neuen Datensatz erstellen. A) Welche Labels werden analysiert/verwendet?  
        Wie oft? Ordnen. Entweder man haut gleich untersten 50% raus oder besser -> B) denn vibrant, high contrast ist kein Stil. B)Insofern Labels mit anderer Stilliste vergleichen -> dann fällts raus MORITZ
  3. Welche Methoden
     1. Erklären. Wie man die anwenden möchte um die Zielfragen zu beantworten GEORG
  4. Wie angewendet?
     1. Daten bereinigt....) neue Files erzeugt
     2. Probleme bei Anyalse
     3. Exact Match
     4. Coocurence…
  5. Statistische Relevanz der Ergebnisse
  6. Visualisieren der Ergebnisse
  7. Einordnung der Ergebnisse (Erklärung, da….)
  8. Zusammenfassung

Realted work finden

Paper und artikel

Paper zu diffusion modellen

Paper zur Digitalen Kunst

Paper zur Methodik (unsere Hypothese)

Artikel zur Frage: warum wird Greg so oft erwaehnt

Ich:

Erster stil ist rausgefallen .> anpassen DONE

n/a raushauen nachschauen bei der lsite wie bankys beeple DONE

nochmal nur kurz checken dasa ganze tags verglichen DONE

Ncohamal testen wie abstract abstract art. Digital . nicht art arrt… DONE

Analysieren welche tags / stile am häufigsten. Wie verteilt etc. also die neue liste unterscuhen DONE

Welche stile wurden nicht gefunden? Vllt fehler daten nochmal schauen DONE

Soviel prozent der küsntler haben in der neuen liste digital als style, wie viele impressionisums DONE

Und dann vergleichen mit den mentions aus der liste

Von georg „preprocessing“ und dann die artists mentions parquet und csv genereieren und damit verhgleichen. Da werden dann die digital wahrscheinlich öfters vor. So viel prozent künstelr liste im vergleich zu artist mentions. Mathplotlib balken ggü stellen.

Wie viel künstler haben kein tag unterschieden zu kein match

Wie viele haben 2 tags wie viele 3

Scene

Mentions

Top künstler der tags

Vllt mal noch die spaöte mitnehmen

Quellenkritik abtract kann eine art sein und ist nicht zwangsweise die direkte Stilrichtung

Quelelnkritik tags sind die repräsentativ aber gubt nichts gerade in diese modern richtung greg rutkowski und so

Forschungsfrage 1:

**Do diffusion model users mention the top artists, such as "greg Rutkowski" and "trending on artstation" to increase the quality of the generated image or to get a specific style?**

Topic modeling und dafür topic greg rutkowski haben

**Hypothesis: Diffusion model users mention the top artists mostly to increase the quality of the generated image.**

Vllt mehr auf Stil fkoussieren! und sagen. 50 % passt zu seinem stil topic die anderen sind vllt weil die qualität wollen, weil andere das nuztzen, weil er populär ist… aus andere gründen

Werden die Künstler gennant um deren Stil zu beommen doer die Qualität des Bildes zu erhöhen?

Kann man für den top künsteler oder top 10 künstler machen

Und dann schauen sind die prompts in denen er vorkommt ähnlcih zu seinem stil oder ganz anders und nur um qualität erhöht werden

Black zombies greg rutkowski macht sinn.

Beautiful rose flower greg rutkowski weniger sinn -> mehr um qualität zu erhöhen

Und das vergleichen mit weniger bekannten / unbekannten Künstler

Und exemplarisch schauen wir uns mal die 3 unbekanten künstler an um zu sehen?! Dass da die prompts mehr auf den stil absehen

Unterfragen (können wir in der Datenanlyse oder als Motivation/Einleitung beantworten):

**Do diffusion model users have a bias towards modern artists? Frage 2**

**Do diffusion model users have a bias towards modern art-styles? Frage 3**

Man könnte dann für Forschungsfrage 2 und 3 die angepasste liste von mir nehemne, aber auch noch für die top 10 künstler wie artgerm oder trending on artstation?! Manuell tags hinzufügen und dann ggü stellen und sagen. Dann ist digital nochmal krasser der beliebteste stil/ tag. Wie verändert scih ergebnis

Wie viel prozent der nennungen haben kein tag. Bewertet damit auch etwas die tagliste der google docs. Und etwas auch die wikiliste und das mapping.

„selbst für ein Künsteler wie artgerm der so vertreten ist gobt es kein tag google docs relativiert. Dennoch ist es die beste und vollstädnigset lsite hat zumindest künstler und tags. Sonst gibt es imm interent noch nichts zu den modernen “