**Portfolioopgave 3**

**1) Beskriv (med referencer) hvorfor institutionen har en API tilgang til deres datasæt, hvad det er blevet brugt til og/eller hvad de tænker det kunne bruges til?**

Jeg tror hovedsageligt at SMK’s har valgt at gøre deres samling tilgængelig via deres API da de ønsker at være en del af den digitale udvikling. Ydermere har SMK, I forbindelses med dette tiltag kaldet ”SMK Open”, udtalt at hensigten er at ”åbne for SMK’s værksamling og sætte en væsentlig del af danskernes kulturarv fri.”.[[1]](#footnote-1) Dette er viser også at SMK ønsker at folk skal have fri adgang til, ikke bare den store mængde kunst, men alt det metadata som kunsten bære med sig. Det kan danne grundlag for ikke bare et vælg at nye interessante undersøgelser, men åbner også op for mange nye muligheden indenfor undervisningens verden. Hvis man f.eks. vil undersøge en bestemt tidsperiode i sin undervisning, kan det være interessant at kigge på hvilken slags kunst denne periode producerede eller lign., eller give studerende kreativ frihed til selv at svarer på lignende spørgsmål vedr. vores kulturelle og kunstneriske historie. Så, for at opsummere, har SMK haft flere grunde til at lave en API til deres datasæt. Ikke alene er det et fremstød på deres digitale front, men også en måde at gøre stoffet mere tilgængeligt for flere former for interaktion, videnskabeligt *og* indlæringsmæssigt.

**2) Sammensæt en URL der søger efter kulturinstitutionens elementer (fx genstande, kunstværker, personer, avisartikler, osv.) i APIen (gerne 100+ stk) med output i JSON. Load denne URL i en JSON beautifier og beskriv indholdet.**

URL: <https://api.smk.dk/api/v1/art/search/?keys=kat&qfields=titles&offset=0&rows=100>

Min URL søger starter med at sørge efter nøglen ”kat”, og specificerer derefter at der kun skal søges efter værker hvor denne nøgle er en del af titlen på selve værket. Dette giver mig 313 resultater, inklusiv internationale kunstværker hvor ”cat” er blevet oversat til ”kat”, og derfor også inkluderet. Inkluderet i alle disse resultater er en stor mængde metadata som f.eks. de materialer som værket er blevet lavet med / af, samt produktionsdato, størrelse(r), m.m.

**2) b) Sammenlign metadata med Dublin Core. Er der overensstemmelser?**

I og med Dublin Core efterhånden er en ret universel formulartype til sortering af metadata, bruges deres element-sæt i mange databaser. Der vil være databaser som kræver andre / flere metadata end DC, alt efter hvad formålet er. Iht. SMK er der flere overensstemmelser, såsom dato, filformat, titel, sprog m.m..

**2) c) Er der nogle tal der kan bruges til statistiske beregninger (fx årstål)?**

Ja, uden tvivl. Der er masser af tal som kan inddrages i statistiske beregninger, såsom årstål, billedstørrelser osv. Ud fra disse kan man f.eks. kigge på præcis hvornår i historien vi har flest billeder fra, om billeder er blevet større eller mindre igennem tiden eller hvor mange malerier med katte vi har fra 2. verdenskrig. Måske endnu mere interessant er alle de ting som *ikke* er tal, som kan bruges til statistiske beregninger, som f.eks. de teknikker kunsten er blevet lavet med. Hvad har været den fortrukne teknik når man har malet i det 20. århundrede, og ændrede det sig f.eks. under de 2 verdenskrige?

**2) d) Hvilke slags spørgsmål kunne besvares gennem analyse af dette datasæt? (gener med eksempler)**

Hvor mange forskellige kunstnere har igennem tiden malet billeder som indeholdt katte? Hvilken kunstner, hvis nogen, har malet *flest* billeder af katte? Hvilken teknik har generelt være fortrukken når man har skulle afbillede en kat? Hvornår har kunsten at male kattebilleder været mest populær?

**3) Ved hjælp af Pandas og request, importer nu jeres valgte datasæt.**

*Se python-kode.*

**3) a) Beskriv og rens datasættet**.

Datasættet er opsat sådan at hver række repræsentere et af de værker som vi fandt gennem vores sammensatte URL, altså er der 313 af disse. Hver kolonne repræsenterer et stykke metadata ud fra det givne værk f.eks. kunstnerens fulde navn (’full name’) eller deres nationalitet (’nationality’), hvis denne kendes.

**3) a) i) Hvilke data og datatyper er der?**

Der er 49 forskellige (meta)meta og disse kommer i form af typer af værdier. Først og fremmest er der objects (’techniques’, ’titles’), hvilket udgør størstedelen af dataen, og så er der også boolean værdier (’on\_display’, ’has\_image’) og floats (’number\_of\_parts’, ’image\_size’), som begge kun findes ved henholdsvis 3 og 4 datapunkter.

**3) a) ii) Hvor mange kolonner og rækker?**

313 rækker og 49 kolonner.

**3) a) iii) Er der kolonner i ikke skal bruge?**

Ja, jeg har valgt at fjerne 17 kolonner af forskellige grunde. Først har jeg fjernet både ’created’ og ’modified’ da disse omhandler hvornår værket blev indskrevet / modificeret i databasen, hvilket ikke bliver nødvendigt for vores statistik (og desuden er identisk meget af det). Derefter har jeg fjernet 3 kolonner som omhandler ’notes’ hvilket, for det første, ikke er tilgængeligt for mange af billeder og, for det andet, var dem som *var* tilgængelige, tilstrækkeligt forskellige til at være for svære at bruge ifm. statistik. Derudover har jeg fjernet flere kolonner som enten var helt tomme eller 95%+ tomme.

**4) Lav nu udtræk og beregn på det data i har udvalgt. Hvilke informationer kan det give os om de udvalgte elementer?**

De 32 kolonner vi står tilbage med kan give os en masse informationer omkring de givne billeder. Vi kan f.eks. kigge på, som tidligere nævnt, hvilke teknikker der er blevet brugt til at producere værkerne, de forskellige tegneres nationalitet eller køn eller måske en graf over hvilke år størstedelen af disse værker er blevet erhvervet eller produceret? Flere af disse spørgsmål vil blive besvaret i opgaver 5 & 6.

**5) Brug value\_counts() (**<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.value_counts.html>**) funktionen til at producere en dataframe over det udvalgte data, optælling og procent.**

*Se python-kode.*

**6) Visualisere det udvalgte data med procent i en graf.**

*Se python-kode.*

Reflektering over processen:

Jeg synes der har været både positiver og negativer i arbejdet med SMK’s API. Overordnet har jeg haft flere oplevelser hvor API’en ikke virkede intuitiv at bruge, da metadata enten var opskrevet på en ret upraktisk måde (f.eks. ifm. kunstneres fødselsdage, men også fordi der er en stor del af resultaterne hvor meget af metadata’en ikke er der. Jeg har også oplevet frustration over at ting såsom kunstnerens navn, fødselsdato og nationalitet, alle sammen lå i kategorien ”production”, og at meget ekstra arbejde derfor at påkrævet, hvis disse skal skilles ad. Også ifm. mit forsøg på at lave statistikker over de teknikker som der er blevet brugt til at lave de forskellige værker, har jeg oplevet problemer. For det første før man et output hvor ordet ’technique’ (unødvendigt) er inkluderet og for det andet så er de forskellige teknikker, i tilfælde af mere end én er brugt, alle opdelt med ’technique’ imellem. Jeg synes personligt at API’en kunne nyde godt af at blive lidt strømlinet, og få tilføjet tags som f.eks. ”tema” eller ”stil”, på en måde der gjorde det nemt at opdele. Der er rigtig mange metadata (se f.eks. sektioner som ender på ”notes”) som sagtens kan analyseres enkeltvis, men er tilstrækkeligt forskellige til at det gør det rigtig svært at arbejde med en masse. Når alt det så er sagt, så er det muligt at få mange interessante ting ud af SMK’s API, og alene det at den eksisterer helt gratis, er et rigtig fedt skridt i den rigtige retning.

Kode-mæssigt så har jeg lært en masse brugbart omkring specielt brugen af pandas samt matplotlib. Jeg er også blevet meget mere komfortabel med ar arbejde med API’er generelt, også selvom SMK’s API, for mig personligt i hvert fald, var noget af en mundfuld. Jeg synes også det er interessant at sætte sig og kigge på hvilke metadata det er relevant at tage med videre i sit datasæt, og så-at-sige ”rengøre” det, så det bliver mere fordøjeligt at arbejde med. Dette var ikke mindst tilfældet ved arbejde med SMK’s API, og selvom det ikke gør den store forskel når datasættet er så relativt småt som det var denne kan, kan det virkelig gøre underværker hvis man arbejder med f.eks. big data.

1. <https://www.smk.dk/wp-content/uploads/2018/06/Projektbeskrivelse_SMK-Open.pdf> [↑](#footnote-ref-1)