

Pràctica Visualització de dades

Selecció del conjunt de dades

1) La rellevància del conjunt de dades en llur context. Son dades actuals? Tracten un tema important per algun col·lectiu concret? S'ha tingut en compte la perspectiva de gènere? (15%)

Les dades que he escollit provenen de dos datasets creats pels usuaris de kaggle "Agirlcoding" i "Akhil Ram". Tots dos conjunts han sigut aconseguits mitjançant scraping de la pàgina web <https://nextspaceflight.com/>. El contingut d'aquest conjunt no contempla la perspectiva de gènere ja que conté informació de les missions espacials fetes des de l'any 1957 fins a l'agost del 2020. Les dades estan disponibles als següents enllaços:

- <https://www.kaggle.com/agirlcoding/all-space-missions-from-1957>
- <https://www.kaggle.com/akhilram7/allrocketsfrom1957>

La raó per la que he escollit aquestes dades és per intentar posar en context el per què de la eufòria que actualment s'està tornant a viure per a la exploració espacial i intentar confirmar si realment estem en una nova era daurada de la exploració espacial. Tot i que els datasets només arribin fins a l'agost del 2020, potser podrem apreciar els primers passos d'objectius futurs tan impressionants com la tornada a la Lluna al 2024 i la colonització de Mart al 2040.

2) La complexitat (mida, variables disponibles, tipus de dades, etc.). Té de l'ordre de centenars o milers de registres? Té de l'ordre de desenes de variables? Combina dades categòriques i quantitatives? Eviteu els conjunts excessivament simples. (30%)

Les dades de tots dos datasets es poden descarregar en format csv.

El primer arxiu conté 4324 llançaments amb 9 variables que els descriuen:

- **2 columnes numèriques** com a identificadors del registre
- **Company Name:** nom de la companyia que va fer el llançament
- **Location:** l'adreça des d'on es fa el llançament.
- **Datum:** data i hora del llançament
- **Detail:** model del coet i la seva càrrega separats per "|".
- **Status Rocket:** variable categòrica amb l'estat actual dels coets ("StatusRetired" i "StatusActive")
- **Rocket:** variable continua amb el cost de la missió en milions de dòlars
- **Status Mission:** variable categòrica amb quatre possibilitats que representa el resultat de la missió ("Success", "Failure", "Partial Failure" i "Prelaunch Failure")

El segon arxiu conté 416 models de coets i 14 variables:

- **S.no:** número identificador del registre
- **Name:** nom del model
- **Cmp:** nom de la companyia que crea el vehicle
- **Wiki:** enllaç amb informació del vehicle a wikipedia

- **Status:** variable categòrica amb tres possible estats del vehicle (“Retired”, “Active” i “Planned”)
- **Liftoff Thrust:** variable continua amb la potència del cohet en Kilo Newtons
- **Payload to LEO:** variable continua amb la càrrega en tones del coet en òrbita terrestre baixa (LEO)
- **Stages:** número de fases que té el coet
- **Strap-ons:** nombre de propulsors suplementaris
- **Rocket Height:** variable continua de l’alçada en metres del coet

3) L'originalitat. No repetiu els conjunts de dades clàssics; podeu, però, millorar visualitzacions existents. Hi ha altres visualitzacions basades en aquest conjunt de dades? És una evolució o actualització d'un conjunt anterior? Heu enriquit un conjunt de dades ja existent? (20%)

Tot i que les dades dels llançaments ja han sigut utilitzades prèviament, usaré la seva versió més recent (la 6a, actualitzada fa 8 mesos) juntament amb les característiques dels coets del segon dataset, que pràcticament no ha sigut utilitzat.

D'aquesta manera, espero poder aportar una perspectiva actual de la història de la exploració espacial gràcies als llançaments actualitzats, i fer èmfasis en les diferències que caracteritzen els models de cada companyia en el temps, utilitzant trets com la mida, la potència o el nombre de fases dels coets.

4) Les qüestions que respondreu amb la visualització de dades: tenen en compte els punts anteriors? Estan ben plantejades? Són adequades pel conjunt de dades triat? (35%)

Una primera aproximació al dataset podria ser respondre les següents preguntes:

- Quina és la importància de cada companyia en la carrera espacial? Quina hi ha gastat més diners? Les companyies d'ara són les mateixes que les del 1957?
- Quins països hi han participat? Amb igual importància al llarg del temps?
- Quins models de coets han sigut els més utilitzats? Hi ha alguna relació amb el cost i les característiques dels models?
- Conforme els anys han anat passant, el cost individual de cada llançament i la inversió anual en missions ha anat variant?
- Quina és la vida útil de cada model de coet?
- Han anat canviant les característiques dels models al llarg del temps? Quina és la última tendència en models? Els possibles canvis van lligats a variacions dels preus de les missions?