



Carles M. Bosch Herrera

Storytelling

ENUNCIAT

L'objectiu d'aquesta prova d'avaluació continuada és que l'estudiant sigui capaç de crear una narrativa de dades amb tècniques de visualització, és a dir, una visualització tipus storytelling.

El lliurament d'aquesta PAC està compost per:

Component 1: Una visualització tipus storytelling.

1) Contextualització i Guió Gràfic:

Objectiu i Usuari: Defineix quin missatge desitges comunicar i a qui va dirigit. Storyboarding: Crea un guió gràfic que esbossi com es presentaran les dades. Això inclou el flux d'informació, els gràfics i els elements interactius si n'hi ha.

2) Selecció d'Eines:

Pots optar per eines com Flourish Studio o Infogram per a una solució sense codi, o llibreries de JavaScript, com ara ScrollyTeller o ScrollMagic per a una experiència més personalitzada i programada. Tens moltes altres eines tant gratuïtes com lliures i propietàries.

3) Disseny i Desenvolupament:

- **Narrativa Visual:** dissenya una història amb estructura narrativa, com ara inici, nus, desenllaç, i com les dades poden generar tensió o resoldre conflictes per mantenir l'interès.
- **Visualització de Dades:** Tria les tècniques que representin millor les teves dades i elimina qualsevol element visual que no aportï claredat i valor. Revisa el treballat a la PAC anterior per associar dades amb representació gràfica.
- **Estètica:** Organitza visualment els elements de manera equilibrada i atractiva, usant colors, tipografies i maquetes (layouts) que reforcin el missatge principal.

4) Revisió i Ajustaments

Assegureu-vos de revisar la vostra visualització per garantir que la narrativa és coherent i els elements visuals són clars i efectius.

Component 2: Un vídeo presentant la visualització. El vídeo ha de tenir una durada entre 4 i 6 minuts.

1) Guió del Vídeo:

Escriu un guió detallat que cobreixi tots els punts exigits: títol, descripció, origen de dades, eines utilitzades, navegació i anàlisi d'elements visuals, i reflexions finals.

2) Enregistrament i Edició

Grava el teu vídeo assegurant una qualitat d'àudio i visual adequada. Mantingues el vídeo entre 4 i 6 minuts, editant per assegurar claredat i concisió.

3) Contingut del Vídeo

- Introducció: Presenta breument l'objectiu de la visualització.
- Metodologia: Explica d'on provenen les dades i amb quines eines es va fer la visualització.
- Demostració: Guia als espectadors a través de la navegació i els elements interactius de la visualització.
- Anàlisi Visual: Discuteix els tipus de gràfics utilitzats, la paleta de colors, la tipografia i com aquests elements ajuden a explicar la història.
- Conclusions: Reflexiona sobre allò que la teva visualització aporta a l'enteniment del tema i com capta l'atenció de l'usuari.

4) Revisió Final

Revisa el vídeo final per assegurar-te que tots els components estan correctament integrats i que es respecta el temps estipulat.

Recursos Addicionals

- Exemples de Storytelling amb Dades:
 - 10 Exemples de Narració de Dades
 - 20 Millors Exemples de Narració de Dades (2022)
- Tutorials i Guies: Com transformar insights d'un quadre de comandes en una narració.

Per a l'elaboració de la visualització us recomanem seguir els següents passos:

- ☐ Pas 1: Entendre el context: Articular la idea principal que es vol comunicar, definir el públic objectiu i crear un guió gràfic.
- ☐ Pas 2: Narrativa: Pensar en els components de la història (inici, nus, desenllaç) i pensar on i com s'hi encaixen les dades? Com poden contribuir la tensió i el conflicte a capturar i mantenir l'atenció del lector?
- ☐ Pas 3: Tria el o els gràfics adequats per explicar les dades.
- ☐ Pas 4: Eliminar el desordre. Hi ha sobers o elements que no aporten valor? Identificar els elements innecessaris i eliminar-los de la visualització.
- ☐ Pas 5: Pensar com un dissenyador. Pensar en com organitzar els elements per crear una estructura visual equilibrada.
- ☐ Pas 6: Decidir: Títols, subtítol, i textos explicatius i d'ajuda...

Rúbrica del vídeo: el vídeo ha de contenir almenys els apartats següents (marcats en groc)

RESOLUCIÓ

1. [10%] Títol i descripció de la visualització creada.

Bona dia. El meu nom és Carles M. Bosch Herrera. Actualment estic cursant el Màster en Ciència de Dades de la UOC i aquest semestre curso entre d'altres l'assignatura de Visualització de Dades.

En aquesta PAC he fet l'**estudi de quina és la situació dels salaris en el sector de les IT en Alemanya**, relacionant diversos atributs com edat, gènere, salari, nivell dintre de l'empresa, tipus i tamany de l'empresa i crearé un **storytelling a partir del dataset utilitzat**.

En aquesta PAC3 m'he centrat en el conjunt d'Alemanya a nivell general i després en la pràctica ampliaré el tema als 6 Länders més poblats d'Alemanya amb la qual cosa ens permetrà tenir una visió general de l'estat de les empreses d'IT a aquest país.

2. [10%] Origen i llicència de les dades.

L'**origen del conjunt de dades** s'ha obtingut de Kaggle. Des de l'any 2015, com ja vaig avançar a la pràctica 1, es realitza una enquesta salarial anònima entre especialistes europeus en el camp de les IT, amb un enfocament centrat a Alemanya. S'utilitzen enquestes a diferents empreses i treballadors d'aquestes que ofereixen voluntaris per participar a aquestes enquestes.

Les dades apareixen a Kaggle a disposició del públic en general i els conjunts de dades contenen informació rica sobre els patrons salarials entre els professionals de les IT a diferents regions de la UE (en aquest cas a Alemanya) i ofereix molta informació que considero que són força rellevants.

A l'informe de l'estat dels salaris al camp de les IT, es proporcionen dades de 2018 a 2020. (<https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>).

El dataset combina tant dades categòriques (ciutat, regió, gènere, nivell a l'empresa,...) com dades quantitatives (edat, salari actual, tamany de l'empresa,...).

La **llicència** per aquest tipus de dades és de domini públic: **CC0:Public Domain**

3. [10%] Eina amb la qual s'ha fet. Característiques d'aquesta eina.

L'eina que he utilitzat per fer aquesta PAC ha estat **Tableau 2024.1.0**, que és una plataforma de anàlisi visual que transforma la manera en què utilitzem les dades per resoldre problemes.



Aquesta eina permet aprofitar al màxim les dades. Algunes de les seves característiques i avantatges que presenta aquesta eina són:

1. **Facilita la exploració i administració de dades:** Tableau facilita l'exploració i la gestió de les dades. Tableau permet connectar-se a diferents fonts de dades, com ara bases de dades, fitxers Excel, fitxers CSV i altres sistemes, i explorar-les de manera intuïtiva.
2. **Permet visualitzacions interactives:** La tecnologia central de Tableau, anomenada VizQL, permet expressar les dades visualment. Això significa que es poden arrossegat i deixar anar elements per crear gràfics, taules i altres visualitzacions sense necessitat de coneixements tècnics massa avançats.
3. **Permet el descobriment d'informació i compartició de manera ràpida:** Tableau ens permet descobrir informació de manera ràpida i compartir-la amb altres usuaris. Això és fonamental per generar canvis significatius en els negocis.
4. **És una plataforma flexible i segura:** Tableau és una plataforma integral que ofereix seguretat i flexibilitat. Tableau s'adapta fàcilment a les teves necessitats de diferents tipus de professionals
5. **Permet extreure informació de les dades emprades i crear un storytelling** per explicar històries de la presentació per explicar al públic en general. Un storytelling és un procés narratiu i estructurat que s'utilitza per transmetre un missatge a través de successos per arribar a una audiència específica.

En resum, podem dir que Tableau ajuda a veure i comprendre les dades, i a generar canvis significatius amb elles.

4. [25%] Presentació de la navegació / animació de la visualització creada.

Inicialment he fet una **visió general de l'estat de les IT a Alemanya a partir de les dades** Es mostren els 6 gràfics implicats en aquesta història a partir dels quals es pot treure informació de les dades proporcionades. Mitjançant els scrolls de la dreta es poden modificar els valors dels intervals dels tres gràfics de barres. Els tipus de gràfics que es mostren són el mapa amb les 16 regions o Länders d'Alemanya (també mostrades a la dreta), un diagrama de sectors, un mapa d'arbre (mal anomenat gràfic de "capses") i tres gràfics de barres que mostren respectivament les distribucions per edat, salari i tamany de l'empresa d'IT.

He fet una **visió general de l'estat de les IT a Alemanya des del punt de la perspectiva de gènere**. A partir de l'anàlisi des de la perspectiva de les treballadores del sector de les IT de sexe femení i de les dades mostrades es comprova que hi ha només un 14 % de dones que treballa en aquest sector de les IT a Alemanya. Dels registres estudiats veiem que no totes les regions o Länders presenten treballadores en aquest sector, cosa que com s'ha vist abans no passa amb el sexe contrari.

També s'ha fet un **diagrama en arbre per representar el nivell a l'empresa (sènior, middle o junior)**.

Finalment s'han fet **3 diagrames de barres per veure la distribució d'edat, salari i tamany de l'empresa**. Mitjançant scrolls es permet variar els intervals dels gràfics d'edat, salari i tamany de l'empresa

A continuació es mostra una taula resum amb els valors dels percentatges i les conclusions a les que he arribat per cada un dels atributs analitzats a partir dels 661 registres d'aquest estudi:

RESUM GENERAL			
	Total Alemanya	Homes	Dones
Regions (total 16 Länders)	8 regions tenen els dos gèneres	4 regions tenen un 100% de gènere masculí	4 regions tenen un 100% de gènere femení
Gènere (Hi ha més homes que dones)	Proporció Homes-Dones 6:1	El 86 % són homes	El 14 % són dones
Nivell a l'empresa (majoritàriament tenen nivell sènior)	El 65.96 % dels treballadors són sèniors	El 71.60 % dels homes són sèniors	El 51.06 % de les dones són sèniors
Edat (majoritàriament a la franja dels 30 anys)	El 61.82 % estan majoritàriament a la franja dels 30 anys	El 62.19 % % estan majoritàriament a la franja dels 30 anys	El 59.57 % estan majoritàriament a la franja dels 30 anys
Salari (Hi ha un sou més baix per les dones, 10 K menys)	El 16.04 % dels treballadors tenen majoritàriament un sou de 65 K	El 17.28 % dels homes tenen majoritàriament un sou de 65 K	El 16.04 % de les dones tenen majoritàriament un sou de 55 K
Tamany de l'empresa (majoritàriament les empreses tenen al voltant de 1000 treballadors)	El 39.49 % són empreses d'uns 1000 treballadors	El 38.80 % dels homes estan en empreses d'uns 1000 treballadors	El 43.62 % de les dones estan en empreses d'uns 1000 treballadors

Comento una mica aquesta taula i les **conclusions** que es poden inferir de l'estudi fet i que es mostren a la taula anterior.

S'ha fet la **distribució per gènere** en el sector de les IT a Alemanya. S'ha utilitzat un diagrama de sectors mostra clarament que hi ha una major proporció d'homes (86%) que de dones (14%). A Alemanya, de la població total estudiada de 661 registres hi ha 567 homes i 94 dones que és aproximadament una relació de 6 homes per cada dona.

Conclusió: El gènere masculí és el més habitual.

La **distribució per edat** en el sector de les IT a Alemanya s'ha fet mitjançant un diagrama de barres. Fent la distribució de les edats dels 661 registres estudiats comprovem que la distribució que presenta un percentatge major és la que va de 20 a 29 anys amb 408 individus que representen el 61.82 % dels individus enquestats. Mitjançant un scroll es permet variar l'interval dels grups d'edat i prendre els valors d'interval de 10, 20, 30, 40 i 50.

Conclusió: L'edat mitja és d'uns 30 anys pels 2 sexes.

La **distribució del nivell dins l'empresa** dels treballadors en el sector de les IT a Alemanya. S'ha plantejat mitjançant un gràfic de mapa d'arbre. Un gràfic de mapa d'arbre representa una espècie de gràfic amb "capses" on es presenta el nivell que s'ocupa a l'empresa (junior, middle o senior). Als mapes d'arbre l'àrea mostrada per cada "capsa" és proporcional a la quantitat de treballadors de cada nivell. En aquest cas la majoria de treballadors enquestats són del nivell sènior que representen 436 treballadors dels 661 registres estudiats, mentre que se nivell middle tenim 188 treballadors i del nivell junior només un 5.6 % (37 treballadors). Si passem per damunt de cada capsa veiem el total de treballadors i el percentatge respecte al total de cada un dels tres nivells.

Conclusió: El nivell a l'empresa més comú pels dos sexes és el nivell sènior.

Per a la **distribució del salari en el sector de les IT** he tornat a plantejar un diagrama de barres. La conclusió que es pot treure és que la majoria dels 661 enquestats tenen un salari entre 55.000 i 69.000 € i fent grups de salari de 5.000 en 5.000 € es mostra una distribució tipus campana de Gauss presentant un màxim entre 60.000 i 64.000 €. Posant-nos damunt de cada barra podem veure el total de registres i el percentatge corresponent. Mitjançant un scroll es permet variar l'interval dels grups de salari i prendre els valors diferents d'interval entre 5.000 i 195.000 € en intervals de 10.000 en 10.000 €.

Conclusió: La mitjana de sou pels homes es d'uns 65.000 € mentre que per les dones és només de 55000 € (uns 10.000 € menys).

En quan a la **distribució per tamany de l'empresa** en el sector de les IT novament he utilitzat un diagrama de barres on s'observa que la majoria de les empreses majoritàriament tenen al voltant de 1000 treballadors amb un 41.34 % del total de les 661 empreses enquestades. Com en casos anteriors, mitjançant un scroll es permet variar l'interval del tamany de l'empresa considerant el total de treballadors.

Conclusió: Les empreses estudiades tenen una plantilla al voltant dels 1000 treballadors.

Passaré a continuació a respondre les dues darreres qüestions que es demanen a la PAC3.



5. [25%] Anàlisi dels elements visuals usats: tipus de gràfics, interacció, colors, textos...

A partir del mapa complet d'Alemanya (on es poden veure els 16 Länders o regions) i dels 6 gràfics analitzats, es poden extreure característiques en el sector de les IT i treure conclusions a partir de les dades del dataset de Kaggle utilitzat.

Per a aquesta anàlisi he utilitzat diferents elements visuals i diferents tipus de gràfics que permeten la interactivitat entre ells. També s'han utilitzat diferents colors i textos explicatius per donar una visualització més clara a l'audiència.

Anàlisi dels elements visuals emprats:

1. Aquesta visualització ha permès, a partir de les dades, mitjançant diferents tipus de gràfics, **copsar quin és l'estat de les registres estudiats en el sector de les IT a Alemanya.**
2. S'han fet servir **6 gràfics interactius**: un del tipus Mapa (utilitzant rols geogràfics, un diagrama de sectors, un diagrama en arbre i tres diagrames de barres per veure les distribucions d'edat, salari i tipus d'empresa. Als gràfics mostrats s'han utilitzat filtres que permeten la interactivitat entre ells i s'han utilitzat **scrolls** per variar els intervals dels diagrama de barres mostrats.
3. Amb diferents **colors i notes de text** juntament amb la interactivitat s'ha intentat captar l'atenció de l'audiència.
4. Tota la informació que es pot extreure ha estat possible a partir dels gràfics i visualitzacions interactives utilitzant **Tableau**.

6. [20%] Reflexions finals sobre què explica i què aporta la visualització creada i quines formes de captar l'atenció de l'usuari s'han utilitzat.

La historia explicada en aquesta visualització permet explicar **quin és l'estat general del sector de les IT a Alemanya considerant diversos aspectes** que vindran donats per les variables utilitzades (incloent la perspectiva de gènere) i aporta força informació fent servir tot un conjunt de gràfics i visualitzacions interactives en Tableau que és una forma de captar l'atenció de l'usuari.

- **Distribució per gènere (diagrama de sectors):** els homes són majoritaris (proporció de 6:1): Homes:86% enfront a les dones:14%
- **Distribució per edat (diagrama de barres):** La majoria de treballadors del sector de les IT, dels registres estudiats, tant per homes com per dones, és jove, entre 20 i 29 anys.
- **Distribució del nivell dins l'empresa (mapa d'arbre):** La majoria de treballadors són del nivell senior (66%), mentre que de nivell middle hi ha un 28.44% i de nivell junior el 5.6%.
- **Distribució del salari (diagrama de barres):** El salari mig es situa entre 55.000 per les dones i uns 65.000 € pel homes.
- **Distribució del tamany de l'empresa (diagrama de barres):** La majoria de les empreses enquestades (41.34 %) tenen al voltant de 1000 treballadors.

A la segona part de la pràctica es continuarà amb aquest estudi, analitzant les 6 regions més poblades d'Alemanya.

WEBGRAFIA I FONTS CONSULTADES:

Com a les PACs anteriors, en aquest document mostro totes les pàgines web i recursos que he consultat per la elaboració d'aquesta document i per l'elaboració d'aquesta part final de l'assignatura corresponent a la PAC3 i a la pràctica..

DATASETS CONSULTATS

IT Salary Survey EU 2020.csv
IT Salary Survey EU 2019.csv
IT Salary Survey EU 2018.csv

WEBGRAFIA CONSULTADA

<https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>
<https://de.wikipedia.org/wiki/Deutschland>
https://es.wikipedia.org/wiki/Alemania#Organizaci%C3%B3n_territorial
<https://querida-alemania.com/organizacion-territorial/>
<https://querida-alemania.com/organizacion-politica/>
<https://es.wikipedia.org/wiki/Alemania>
<https://de.wikipedia.org/wiki/Berlin>
<https://data.world/datasets/germany>
https://data.europa.eu/data/datasets/https-www-genesis-destatis-de-genesis-downloads-00-tables-12411-0009_00?locale=ca
https://data.europa.eu/data/datasets/https-www-regionalstatistik-de-genesisws-downloader-00-tables-ai010-1_00?locale=ca
<https://data.europa.eu/data/datasets/d8460e47-358f-45c8-8256-daf3e4d519ca?locale=ca>
<https://data.europa.eu/data/datasets/ba3ace60-4d2e-4694-47f7-326b01ab6245?locale=ca>
<https://cursos.frogamesformacion.com/courses/tableau-basic>
<https://cursos.frogamesformacion.com/courses/tableau-avanzado>
<https://knoema.es/atlas>
<https://knoema.es/dflijuc/germany-regional-dataset>
<https://knoema.com/atlas/Germany/datasets?mode=amp>
<https://knoema.com/atlas/Germany>
<https://querida-alemania.com/organizacion-territorial/>
<https://querida-alemania.com/organizacion-politica/>
<https://www.superdatascience.com/pages/tableau>
<https://stats.oecd.org/Index.aspx?DataSetCode=RPOP>
<https://www.udemy.com/course/tableau-a-z/>
<https://www.udemy.com/course/tableau-a-z-avanzado/>