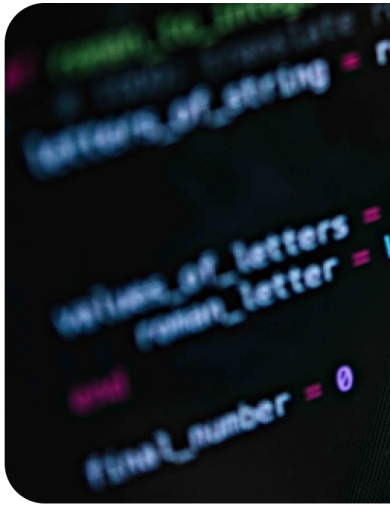
Project Analysis Ideas

Salaries in AI, ML, and Big Data



le wagon

Overall Analysis

Analysis Ideas

Overall Analysis

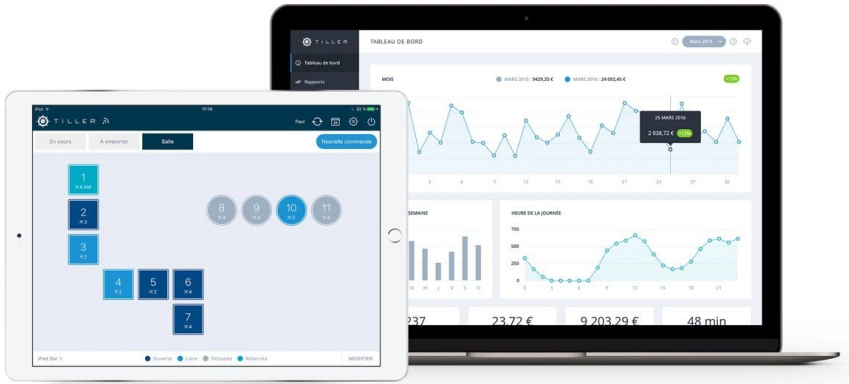
- Work year distribution (which year appears most in the dataset)
  - Pie chart (slices represent each year)
- Remote options distribution (distribution of remote work options among the dataset entries)
  - Pie chart (slices represent each remote option)
- Experience level distribution (distribution of experience levels among the dataset entries)
  - Bar chart (x= experience level, y=experience entry counts)
- Employment type distribution (distribution of employment types among the dataset entries)
  - Bar chart (x= employment type, y=employment entry counts)
- Job title distribution (choose top 10 job title by number of entries)
  - Bar chart (x= count of job title entries, y=top 10 job titles)
- Employee residence distribution (choose top 10 countries of employee residence)
  - Bar chart (x=count of residence countries entries, y=top 10 country names)
  - Map chart (show every entry point in map - density areas?)
- Company location distribution (distribution of company locations among the dataset entries)
  - Bar chart (x=count of company countries entries, y=top 10 country names)
  - Map chart (show every entry point on map - density areas?)

Simple Analysis

Analysis Ideas

Simple Analysis

- Salary variation per experience levels
  - Bar chart (x= experience level, y=salary)
  - Box chart (x= experience level, y=salary)
- Salary variation per employment type
  - Bar chart (x= employment type, y=salary)
  - Box chart (x= employment type, y=salary)
- Salary variation per remote options
  - Bar chart (x= remote options, y=salary)
  - Box chart (x= remote options, y=salary)
- Salary variation per company size
  - Bar chart (x= company size, y=salary)
  - Box chart (x= company size, y=salary)



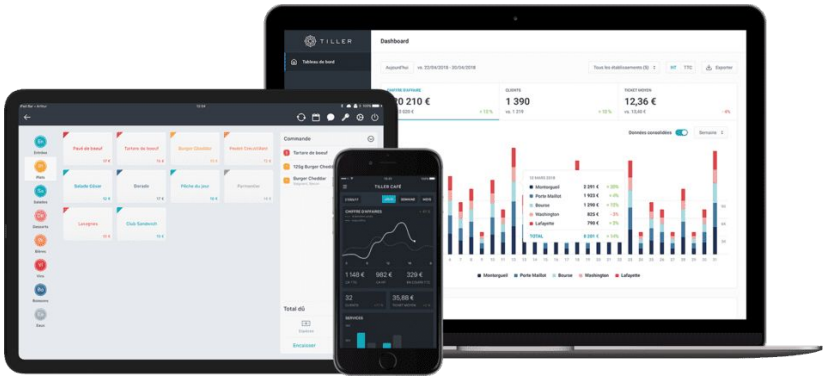
le wagon

In Depht Analysis

Analysis Ideas

In Depth Analysis

- Salary based on experience level by employment type
  - Bar chart (x= experience level, y=salary, color/hue=employment type)
- Salary based on experience level by company size
  - Bar chart (x= experience level, y=salary, color/hue=company size)
  - Box chart (x= experience level, y=salary, color/hue=company size)
- Salary based on experience level by remote options
  - Bar chart (x= experience level, y=salary, color/hue=remote option)
- Salary fluctuations over the years
  - Line chart (x= year, y=salary)
- Salary variations over the years based on remote options
  - Bar chart (x= year, y=salary, color/hue=remote option)
- Salary based in remote options and employment type
  - Bar chart (x=employment type, y=salary, color/hue=remote option)
- Salary based in company size and experience level
  - Histogram chart (x=company size y=salary, color/hue=experience level)
- Salary based on experience level, employment type and remote option
  - 1. Bar chart (x= experience level, y=salary, color/hue=employment type) **for full remote**
  - 1. Bar chart (x= experience level, y=salary, color/hue=employment type) **for hybrid**
  - 1. Bar chart (x= experience level, y=salary, color/hue=employment type) **for no remote**

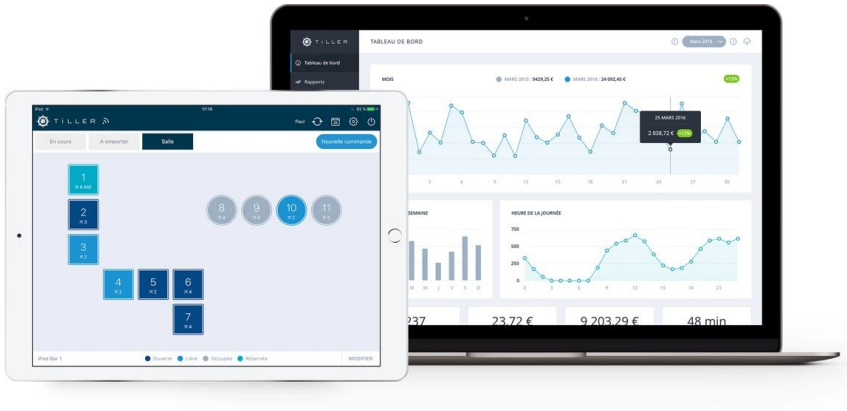


le wagon

Analysis Ideas

In Depth Analysis

- Salary based on employee residence by experience level
  - Bar chart (x= employee country, y=salary, color/hue=experience level)
- Salary based on company location by experience level
  - Bar chart (x= company country, y=salary, color/hue=experience level)
- Compare salaries of the company location vs employee residence
  - Compare previous two bar graphs
- Salary trends by experience level and company size
  - Bar chart (x= experience level, y=salary, color/hue=company size)
- Salary evolution over time among top 10 job titles
  - Bar chart (x=year, y=salary, color/hue=job title(top 10))
- Salary based on remote option by job title
  - Histogram chart (x=remote option, y=salary, color/hue=job title)
- Salary trend by job title across experience levels
  - 1. Bar chart (x= job title, y=salary) **for entry level**
  - 1. Bar chart (x= job title, y=salary)**for mid level**
  - 1. Bar chart (x= job title, y=salary) **for senior level**
  - 1. Bar chart (x= job title, y=salary) **for executive level**

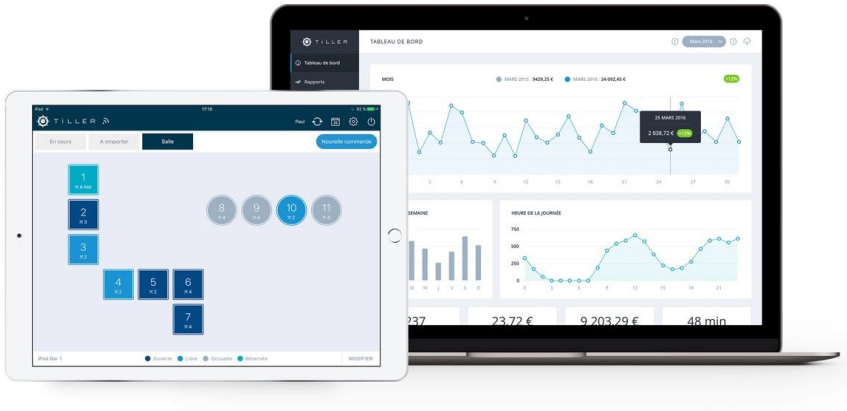


le wagon

Analysis Ideas

In Depth Analysis Notes

- Base all the salary measures in the **salary_in_usd** column (USD currency) in order to guaranteeing a balanced and equal analysis of all the data
- Whenever makes sense add a graph solemnly for **data analyst**
- Make **ML previsions** regarding interesting insights we found
- Procurar tabelas para adicionar data sobre backgounds dos entry-levels (como se graduaram - bootcamp, bachelor, self-learning...)
- In graphs that makes sense do a top 10 and bottom 10
- Whenever possible, include Portugal data
- Make two profiles, for e.g from a data analyst student that wants to work in remote and want to understand where to send his resume to, and a senior BI engineer searching a job in a large company in a foreign country
- In case we don't have time we can say that "in the next year we will have these features added to the new version"
- Start the presentation explaining the dataset, referring the fields with the biggest percentage (e.g. US has the largest percentage, Medium size also, and Full time also)
- Explain what is a Small, Medium and Large company
- Idea: cross our data with tables that have values solemnly to data scientists, data analysts... to see what are the insights we take with more variety of data, because our sample has a giant amount of entries from US, full-time work mode and medium size companies
- **Chose a color palette to use in the logo and charts (and possibly a shirt)**
- **Por os valores SEMPRE em average por causa da disparidade de numero de entries na nossa tabel**



le wagon

- Ver sempre qual a quantidade de valores (entries) que temos da metrica que estamos a avaliar para perceber se a average nao esta a ser influenciada e a criar um bias.
  Porque se tivermos menos entries de x metrica o average vai sempre ser maior que o de uma que temos muita data.
- Explicar que temos 3 tipos de case (profiles que procuram informações diferentes na nossa data)
- Pegar na nossa ideia de salario de alguem que vai a uma entrevista de emprego e quer responder quanto espera receber a fazer aquele trabalho
- Do a disclaimer at the end of the presentation a dizer que a maioria dos nossos dados sao mais indicados para os US porque temos mais data de la
- Comparar os data analysts com a media do cluster de jobs de IT, em relação a campos que achamos pertinentes.

- **IDEAIS DOS GRAFICOS**
  No primeiro grafico calcular a percentagem de full time de cada experience level para podermos ver a variacao entre junior e senior, ou intermediate e executive... avaliar o gap
- Nao por o full time, contract... por 90 e tal % da data ser full time, o que faz com que os dados estejam viciados
- Uma ideia é separar os US do resto do mundo, uma vez que temos a maioria da data vinda dos US
- Usar o tree graph para mostrar as quantidade de data que temos de cada pais e podemos tambem usar esse grafico como um filtro para filtrar a outra parte dos graficos. Carregar na parte do US para aparecer dados do US nos outros graficos todos


le wagon