

Genome - O que é ou o que faz

Genome, é uma startup focada em catalisar o crescimento de ecossistemas de inovação e garantir que o máximo de cidades possam fazer parte e contribuir para os movimentos de desenvolvimento tecnológico.

Escolhemos esse tema porque

O mercado de patentes é um tema altamente rico em dados e capaz de fornecer diversos insights, possibilitando antecipar possíveis inovações nas tecnologias já existentes e outras tendências que podem vir a surgir nos próximos anos.

Nossas bases de dados são composta por

Dados geográficos a respeito da concessão de patentes(país, cidade e ecossistema), classificação dos segmentos tecnológicos que cada patente faz parte, dados temporais a respeito das concessões (ano e data em que foram concedidas).

Com esse desafio esperamos

Obter maior entendimento e extrair insights a respeito do cenário mundial de pesquisa, desenvolvimento de patentes e direcionando esforços da Genome para atuação em ecossistemas e subsetores com maior taxa de crescimento.

Quais são as suas hipóteses sobre o tema?

- Hoje o EUA é um polo forte com relação a concessão de patentes, mas o continente asiático vem despontando nos últimos anos na obtenção de novas patentes e aparenta ser um ecossistema muito promissor para a atuação da Genome.

Quais perguntas elas levantam?

- 1) Os ecossistemas da Ásia estão se desenvolvendo mais rapidamente do que os EUA ao longo dos anos?
- 2) O perfil de desenvolvimento de patentes dos EUA é diferente do perfil dos ecossistemas asiáticos?
- 3) Que segmentos de patentes vem despertando mais interesse dos setores de pesquisa e inovação?

Qual(is) delas os seus dados podem responder?

- Com os dados que nós temos atualmente, acreditamos ser possível responder as perguntas: 1,2,3

O que e como isso será feito?

- Através de análises da taxa de crescimento de concessão de patente por ano, país, ecossistema e analisando as categorias das patentes concedidas.

Qualidade dos Dados

```
Data columns (total 17 columns):
```

#	Column	Non-Null	Count	Dtype
0	patent_id	320267	non-null	int64
1	name_first	320267	non-null	object
2	name_last	320209	non-null	object
3	city	320249	non-null	object
4	country	320266	non-null	object
5	grant_date	320267	non-null	object
6	grant_year	320267	non-null	int64
7	app_date	320266	non-null	object
8	app_year	320266	non-null	float64
9	class_distinct_count	319067	non-null	float64
10	class_concat	319067	non-null	object
11	class_IPC_distinct_count	319067	non-null	float64
12	class_IPC_concat	318821	non-null	object
13	CountryName	320247	non-null	object
14	CityCountry	320266	non-null	object
15	Ecosystem	237694	non-null	object
16	GoogleCity	241834	non-null	object

dtypes: float64(3), int64(2), object(12)
memory usage: 41.5+ MB

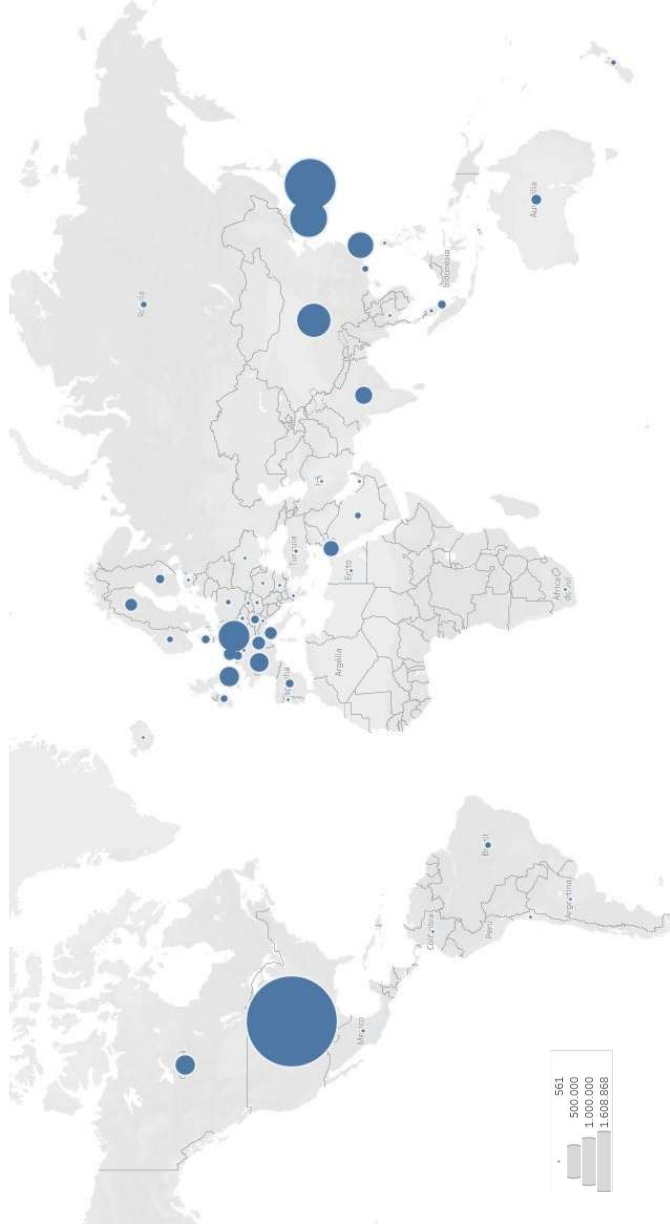
A principal fonte de dados que recebemos para realizar as análises neste projeto foi a base “raw_patentes.csv”. Em um primeiro momento, conduzimos uma análise exploratória para entender que tipos de dados estavam presentes neste dataset e avaliar a qualidade destes dados.

Na imagem ao lado podemos ver que a maioria das colunas está totalmente preenchida, apenas as últimas duas colunas, “Ecosystem” e “GoogleCity” apresentam uma maior número de linhas em branco.

Isso acaba sendo um problema porque os dados de ecossistemas são bem importantes para as nossas análises geográficas e, por conta disso, ao longo do projeto estudaremos meios para completar estes dados.

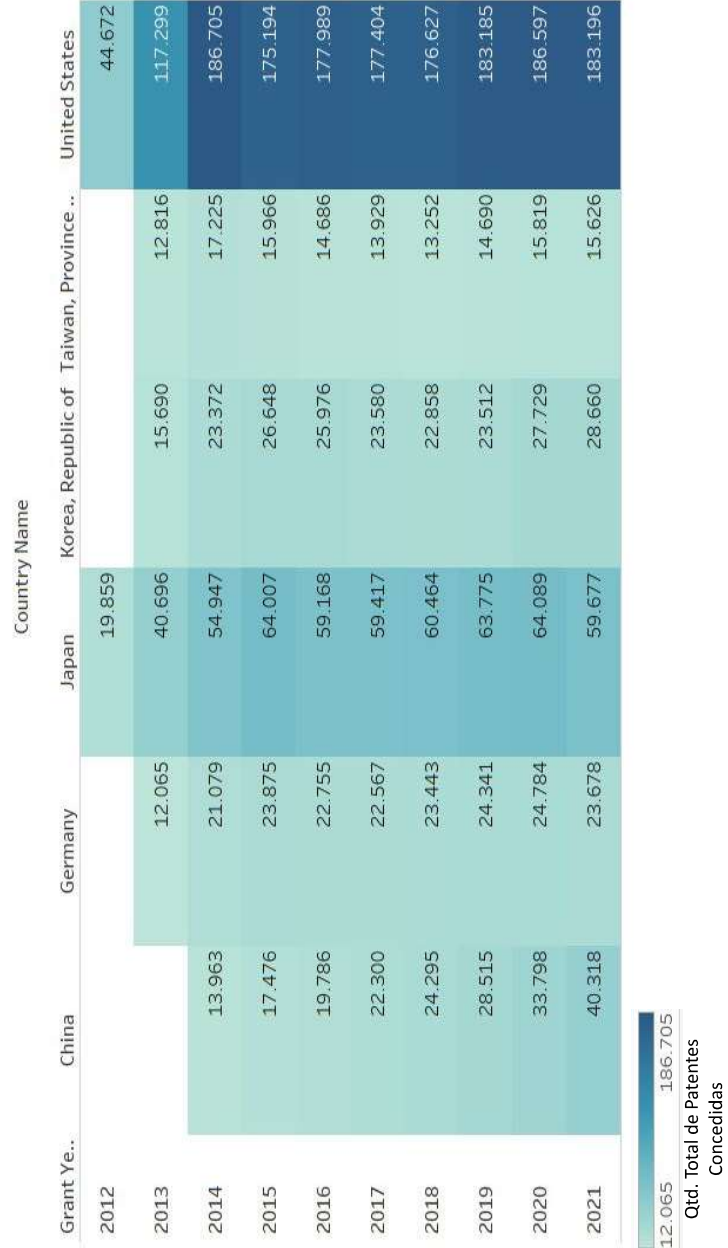


Análises - País



A primeira análise que conduzimos no dataset, após averiguar a qualidade dos dados, foi voltada para estudar quais países possuem o maior volume de patentes concedidas e com isso começar a entender para qual “direção” o eixo tecnológico está voltado no âmbito global.

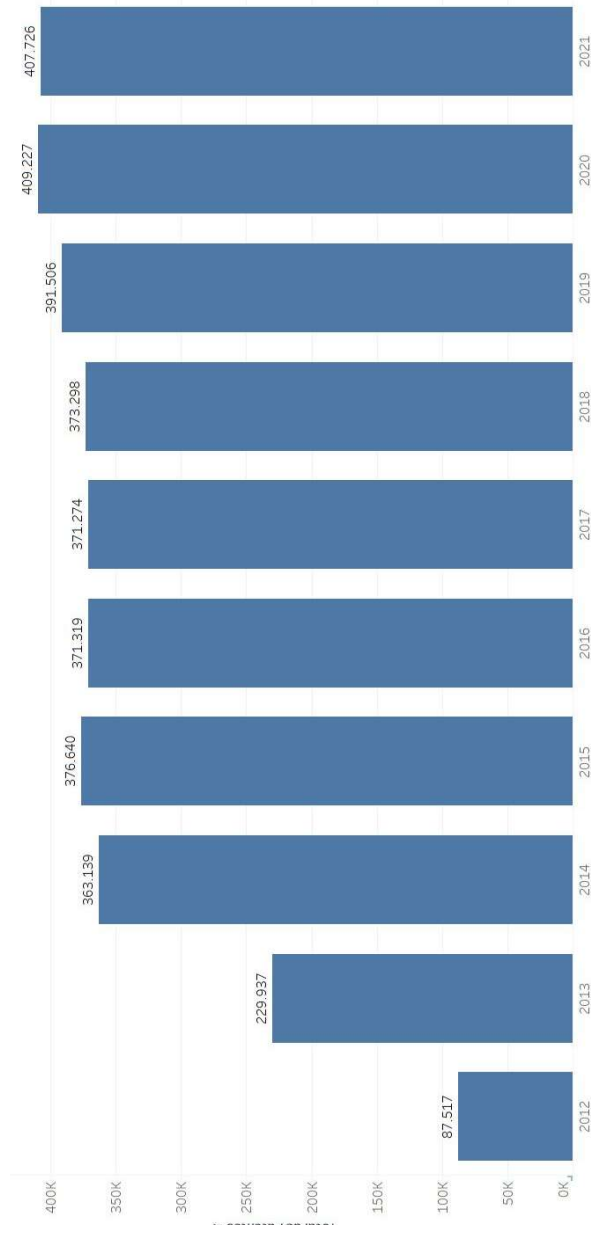
Análises - País



A partir do gráfico anterior, conseguimos encontrar os principais “players” mundiais e o próximo passo foi estudar a evolução temporal desses países em relação às concessões de patentes.

A ideia deste estudo foi entender melhor as tendências de crescimento destes países e analisar quais já possuem um mercado mais consolidado, quais estagnados, quais apresentam queda e, principalmente, quais vem despontando nos últimos anos, o que pode indicar um cenário mais “aquecido” no desenvolvimento de patentes

Patentes - Análises

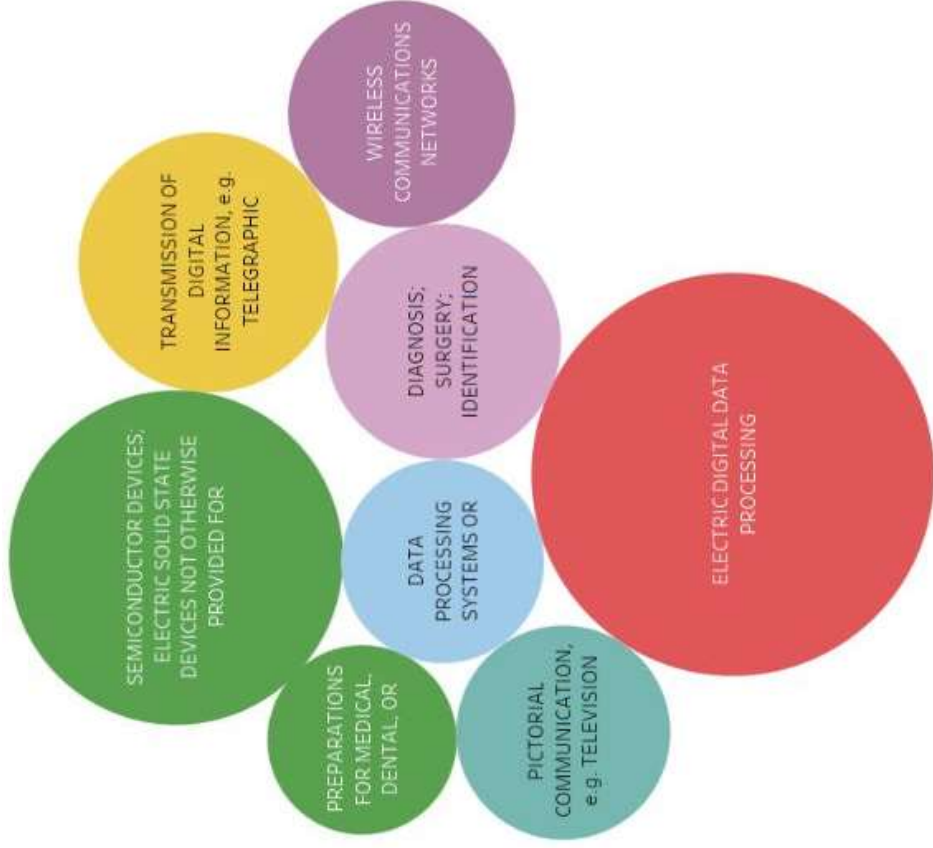


Na imagem ao lado podemos ver a evolução temporal na concessão de patentes por ano. Há um grande salto entre 2012/13 e 2014, provavelmente ocasionado por questões de séries temporais, por se tratar do começo da coleta dos dados.

De 2014 em diante é possível ver um crescimento lento e estável até 2020 e em 2021 uma leve queda.

Por mais que o volume total tenha um comportamento “estável”, é importante analisar a composição destas patentes para entender o contexto tecnológico global e como cada ecossistema performa.

Patentes - Análises



Nesse gráfico visualizamos as classificações de patentes (IPC) de maior destaque. A maior parte delas está associada ao processamento de dados digital (G06F) ou eletricidade (H), com semicondutores (H01L) ou sistemas de transmissão de informação digital (H04L) que inclui transmissão de dados.

O próximo passo seria relacionar o IPC com os subsetores definido pela Genome. Como foi fornecido o resumo de cada patente podemos aplicar machine learning para classificar cada patente em um subsetor.