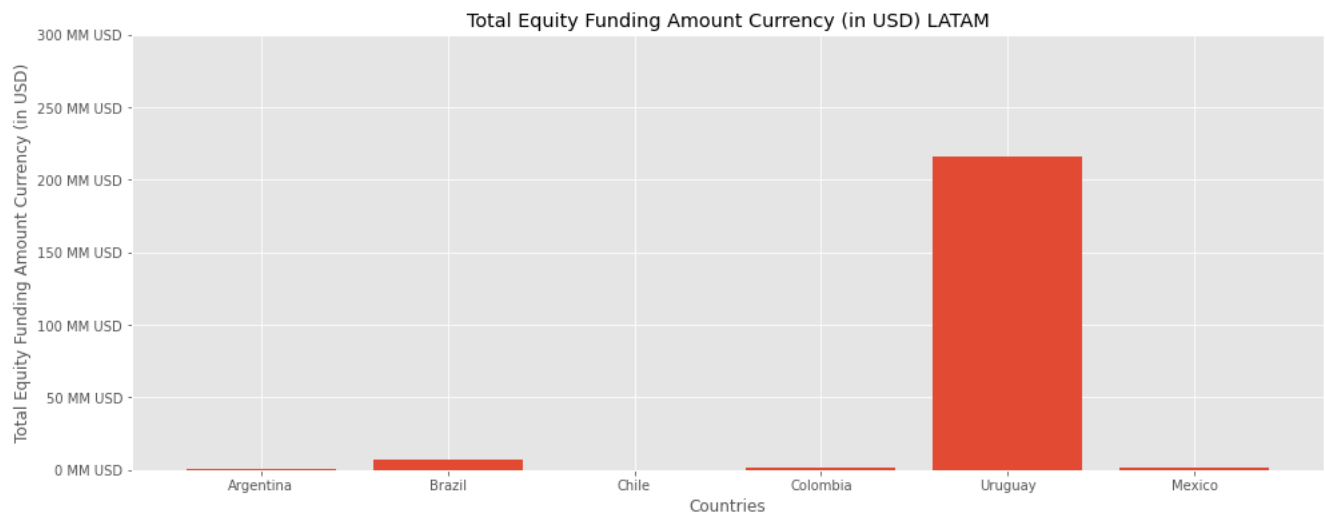


Taller 5

Punto 1. Cuánto capital se ha invertido en LaTAM durante el último año. Desagregue gráficamente por país.

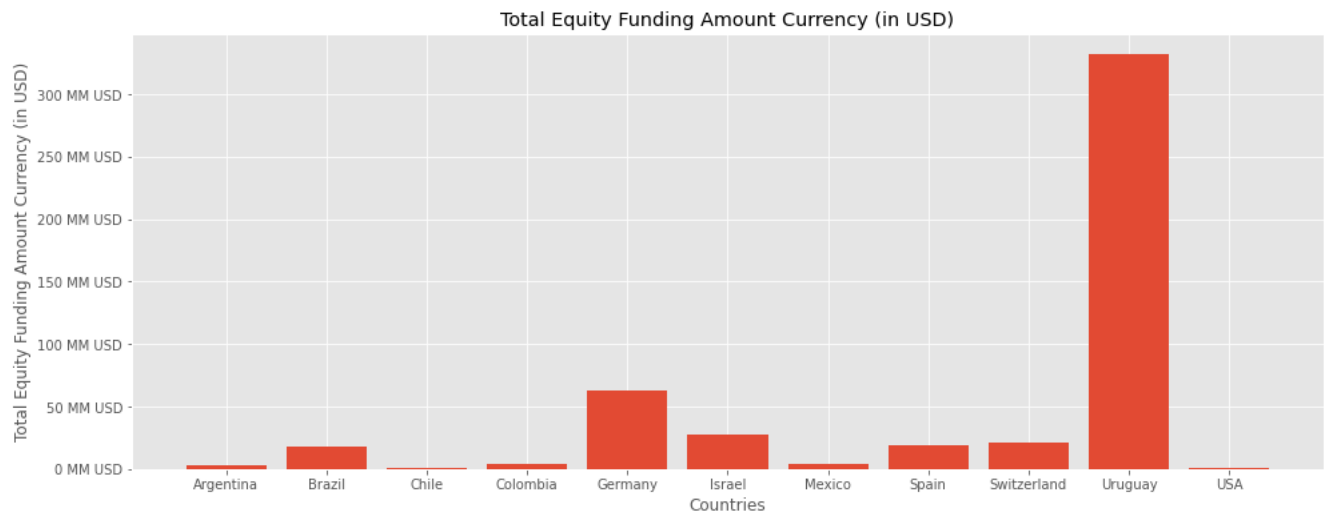
Solución. Filtrando la base de datos, quedándonos solo con el capital recaudado en Latinoamérica, cuya fecha se encuentra entre los años 2020 y 2021, obtenemos la gráfica.



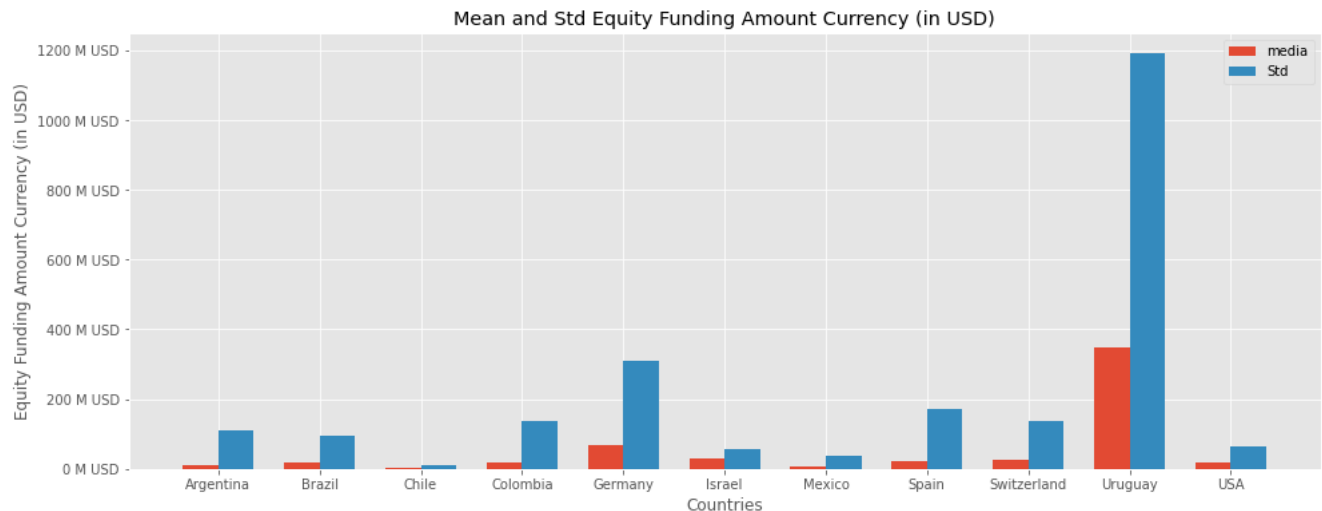
Donde observamos que Uruguay es el país latinoamericano que entre el año 2020 y 2021 obtuvo mayores recaudaciones, Opacando a los demás países, solo sobresaliendo entre estos Brasil.

Punto 2. Haga una comparación entre Colombia con cada uno de los otros países. Analice.

Solución. Como primer acercamiento miraremos el monto total recaudado en generar por cada país.



Donde de nuevo Uruguay sobresale por ser el que más recaudo obtiene, y Colombia siendo uno de los que menos, para comprobar que esto no se debe a un valor atípico observaremos como es la media de el capital recaudado y su desviación.

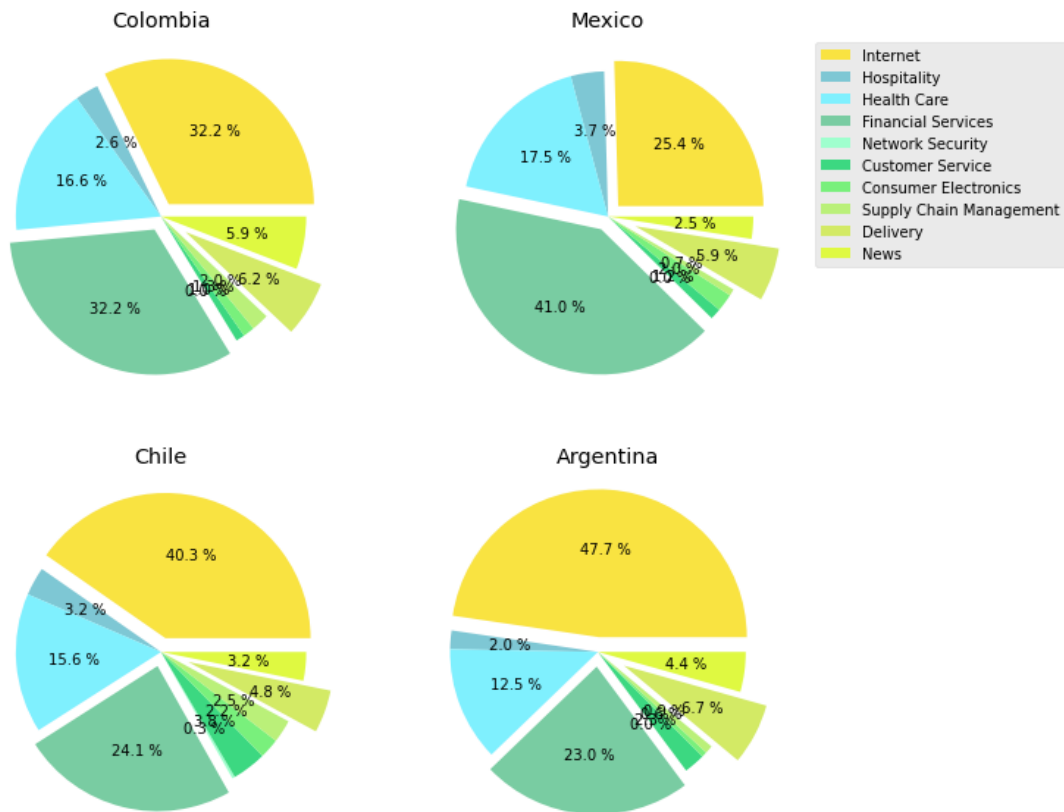


La desviación, la cual es mayor en Uruguay nos dice que aunque es el país que obtiene mayores recaudaciones, la desviación es bastante grande lo que nos sugiere que estos montos se agrupan en pocas empresas y la mayoría obtiene recaudaciones pequeñas, un comportamiento similar lo vemos con Colombia, donde la media no sobresale respecto al resto pero su desviación si.

Realizando un análisis sobre las industrias, las industrias con mayores recaudaciones entre todos los países son

Industries	Number
Internet	993
Hospitality	98
Health Care	976
Financial Services	966
Network Security	95
Customer Service	93
Consumer Electronics	93
Supply Chain Management	92
Delivery	92
News	90

de aquí podemos visualizar su comportamiento en tres países de América latina y Colombia



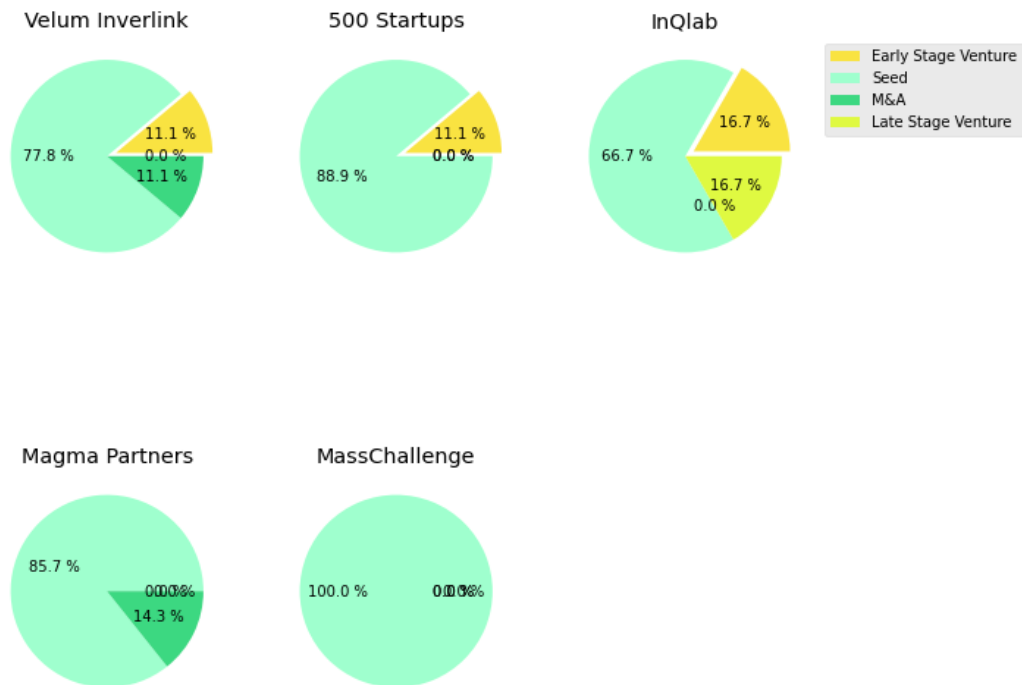
Países como Colombia, Chile y Argentina suelen invertir bastante en internet y Mexico en Financial Services, siendo el segundo con mayores recaudaciones en estos cuatro.

Punto 3. ¿Cuáles son los fondos que más invierten en Colombia? Haga un análisis descriptivo de cada uno de ellos.

Solución. Los inversores que más invierten en Colombia son

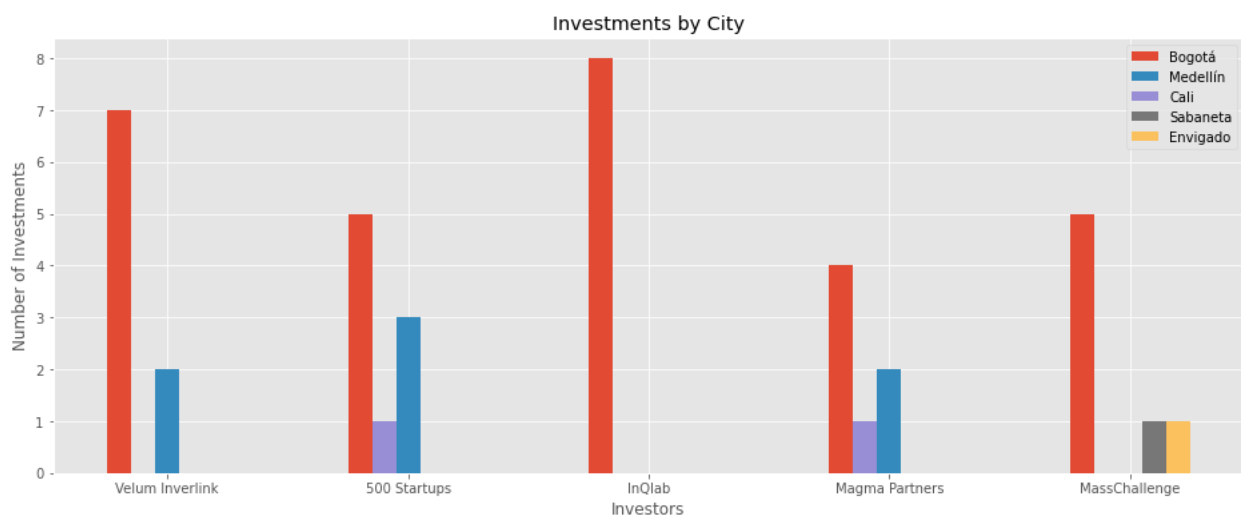
Investors	Number
Velum Inverlink	9
500 Startups	9
InQlab	8
Magma Partners	7
MassChallenge	7

Una primera idea sobre sus intereses de inversión son si están interesados en las empresas pequeñas o en las ya avanzadas



Donde observamos que la mayoría de inversiones se encuentran en las empresas pequeñas "Seed".

Por tanto centraremos nuestra vista a las ciudades donde más se concentran dichos inversores



Pero como era de esperarse, esta se concentran en Bogotá y Medellín, veamos entonces sus intereses en los sectores industriales.

Industries Velum Inverlink	Number
E-Commerce	5
Retail	2
Home Decor	1
Subscription Service	1
Communities	1

Industries 500 Startups	Number
FinTech	2
Software	2
Travel	2
Health Care	2
Financial Services	2

Industries InQlab	Number
Financial Services	3
FinTech	2
Information Technology	2
Software	2
Customer Service	1

Industries Magma Partners	Number
Financial Services	2
FinTech	2
Artificial Intelligence	2
Finance	2
Developer Platform	2

Industries MassChallenge	Number
Accounting	1
Financial Services	1
Internet	1
Education	1
Machinery Manufacturing	1

Todos estos inversores suelen centrar en invertir en selecciones diferentes, MassChallenge no posee una preferencia por algún sector a diferencia de Velum Inverlink que se destaca por apoyar a la industria de "E-Commerce", por otra parte aparece en común a los otros tres inversores el sector de "Financial Services".

Punto 4. Muestre gráficamente los exits de capital privado en Colombia por deal size.

Solución. La base de datos no nos proporciona información relevante en cuanto a los exits de capital privado, ya que al filtrar sólo nos encontramos con una coincidencia.

Organization Name	Number of Exits
Veronorte	1

Punto 5. Realizar una regresión logística para determinar las características que hacen exitosa una startup para obtener inversión. Hacer el análisis correspondiente.

Solución. Por la enorme cantidad de variables que no aportan información alguna para determinar si una startup es exitosa o no, centraremos nuestro estudio en las siguientes variables "Headquarters Location", "CB Rank (Company)", "Estimated Revenue Range", "Last Funding Type", "Number of Employees", "Number of Articles" y "Industry Groups".

Como "Industry Groups" viene empaquetados desglosamos esta variable y creamos nuevas variables correspondiente a cada sector industrial, junto con esto creamos variables dummy para las categoricas como la ciudad y tipo de fondo, obteniendo un total de 144 variables para la regresión logística.

Con ayuda del paquete podemos filtrar de estas 144 variables, cuáles aportan información relevante en la clasificación de los startup.

```
from sklearn import datasets
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression

X = list(Data_regres.columns)[:144]
Y = Data_regres.columns[145]

lr = LogisticRegression()

rfe = RFE(lr,12)
rfe = rfe.fit(Data_regres[X],np.array(Data_regres[Y]).reshape(-1,1))
```

Seleccionamos las variables que el modelo nos dijo que aportaban.

```
X_selec = np.array(X)[rfe.ranking_ < 50]
X_selec
```

y entrenamos nuestro modelo para estas.

```
from sklearn import linear_model

logit_model = linear_model.LogisticRegression()
logit_model.fit(Data_regres[X_selec],np.array(Data_regres[Y]).reshape(-1,1))
```

podemos imprimir las variables que aportan a este modelo.

```
array(['Headquarters Location_Bogotá, Distrito Especial, Colombia',
      'Headquarters Location_Boyacá, Boyaca, Colombia',
      'Headquarters Location_Cali, Valle del Cauca, Colombia',
      'Headquarters Location_Canadá, Cundinamarca, Colombia',
      'Headquarters Location_Cundinamarca, Distrito Especial, Colombia',
      'Headquarters Location_Envigado, Antioquia, Colombia',
      'Headquarters Location_Madrid, Distrito Especial, Colombia',
      'Headquarters Location_Manizales, Caldas, Colombia',
      'Headquarters Location_Perú, Valle del Cauca, Colombia',
      'Estimated Revenue Range_$100M to $500M',
      'Estimated Revenue Range_$10M to $50M',
      'Estimated Revenue Range_$1M to $10M',
      'Estimated Revenue Range_$50M to $100M', 'Last Funding Type_Angel',
      'Last Funding Type_Convertible Note',
      'Last Funding Type_Corporate Round',
      'Last Funding Type_Debt Financing',
      'Last Funding Type_Non-equity Assistance',
      'Last Funding Type_Pre-Seed', 'Last Funding Type_Private Equity',
      'Last Funding Type_Seed', 'Last Funding Type_Series A',
      'Last Funding Type_Series B', 'Last Funding Type_Series D',
      'Last Funding Type_Series F', 'Last Funding Type_Undisclosed',
      'Last Funding Type_Venture - Series Unknown',
      'Commerce and Shopping', 'Information Technology', 'Real Estate',
      'Sales and Marketing', 'Financial Services', 'Mobile', 'Software',
      'Other', 'Professional Services', 'Transportation', 'Hardware',
      'Data and Analytics', 'Apps', 'Community and Lifestyle',
      'Artificial Intelligence', 'Science and Engineering',
      'Privacy and Security', 'Events', 'Travel and Tourism',
      'Advertising', 'Video', 'Health Care', 'Consumer Electronics',
      'Biotechnology', 'Agriculture and Farming', 'Sustainability',
      'Manufacturing', 'Energy', 'Natural Resources', 'Gaming',
      'Clothing and Apparel', 'Design', 'Platforms'], dtype='<U65')
```

y su significanos para la clasificación.

```
logit_model.coef_
```

```
array([[ 0.24848365,  0.8429826 ,  0.21540652, -0.22048297,  0.45247667,
        -0.21826737, -0.4623711 ,  0.65305225, -0.22700138, -0.35124953,
        -0.20802765,  0.20702714,  0.26093287, -0.5079318 ,  0.19340338,
        -0.25626288,  1.02999874, -2.02437128, -0.34137096, -1.23216188,
        -0.23500394,  1.61931452,  1.53516669,  0.71292091,  0.8252619 ,
         0.66888841,  0.52169984,  0.61513507, -0.24435687,  0.29734695,
        -0.73313301,  0.7346034 ,  0.71275629,  0.44567818,  0.55797907,
        -0.43230675,  0.50997338, -0.62230818,  0.72939569, -0.22480144,
        -0.69002887, -0.35459596, -0.29988372,  0.28951745, -0.26813294,
        -0.68758024, -0.48142057, -0.34931599,  0.72481936, -0.43444785,
        -0.38190176,  0.45978927, -0.49249512, -0.42826506, -0.51867025,
        -0.25957198, -0.40756261, -0.22554175, -0.47924617,  0.35773504]])
```

Y por ultimo la precisión de el modelo para determinar si un startup sera exitoso o no.

```
logit_model.score(Data_regres[X_selec],np.array(Data_regres[Y]).reshape(-1,1))
```

```
0.958
```