CAPSTONE PROJECT REPORT – THE BATTLE OF AFRICAN CITIES

FINDING SILICON SAHARA

SYLVESTER EKOW HARRISON
Feb. 17, 2019

Abstract
In Africa, Digital technologies continue to spread rapidly. Digital dividends, the aggregate impact of these technologies, however, has fallen short thus far. This has been one of the key features of the African digital renaissance.
Nonetheless the African continent remains the one of the most promising digital ecosystem markets on the planet today. It is projected that 1 in 4 people worldwide will be African by 2030. As the Region's population soars and more people come online for the first time, businesses and the region's economy overall will become more digital as entrepreneurs, stakeholders and governments adapt and leverage digital transformation for efficiency, better decision-making and solutions better suited to the "next billion".
As our businesses become more digital, a Silicon Valley-like model of inclusion, efficiency and innovation, modified to cater to the African context, must be replicated to effectively maximize the digital dividends. Spencer, Vinodrai, Gertler, & Wolfe, 2009 posits that 'there are measurable benefits for related economic activities to locate in the same metropolitan region'.
In this context, identifying Africa's prime digital innovation hub, city or region is imperative. Such a region must stand out positively with a geographic advantage, favorable economic reforms, and a mature start-up ecosystem to transform Africa's fast-growing tech ecosystem into a large integrated network and allow for the complete digitization of the African economy.
In this project, a k-means clustering algorithm and spatial analytics/location data of the Foursquare API is used to explore and segment Africa's technology ecosystem landscape into economic clusters to identify Africa's prime digital technology hub, city or region, ***Silicon Sahara*** *to lead Africa's digital revolution.*
Keywords – Clustering; k-means algorithm; Foursquare API; location data; technology hubs; digital revolution

---

I. INTRODUCTION

A. Background

For decades, the Silicon Valley model has progressively transformed the way innovation and entrepreneurship is viewed globally. This has inspired technology-minded individuals to form clusters, across various geographical regions, to collectively shape the world with technological innovation that mirrors that of Silicon Valley. From London's 'Silicon Roundabout' to Dubai's 'Silicon Oasis', governments and businesses around the world are keen to brand their local tech industries as the next hotbed of innovation. With young, digitally-savvy populations, African cities are also joining in on the trend: 'Silicon Savannah', 'Silicon Cape' and 'Silicon Lagoon' have emerged as popular monikers for Nairobi, Cape Town, and Lagos.
Home to thousands of technology start-ups, Africa's estimated $50bn tech ecosystem is now a recognized and thriving tech industry and one of the fastest growing technology ecosystems in the world. The goal of technology hubs is not only to offer space and services to start-ups but also to create a large community of stakeholders (partners, start-ups, investors etc.). Successful network effects are often a powerful tool for organic growth to foster businesses and technology ecosystems and spur economic growth. This makes technology hubs and start-ups an essential part of the digital world.
More cities are emerging as regional innovative hubs and there is now a growing battle between some of Africa's largest cities and regions as the leader of Africa's thriving technology ecosystem landscape.

B. Business Problem

Interest in the continents digital advancements is well documented with tech leaders and visionaries including Mark Zuckerberg (Facebook), Sundar Pichai (Google), and Jack Ma (Alibaba) having visited major hubs from Lagos to Nairobi. The past couple of years have also been characterized by the rise of worlds tech giants' activity and interest in the untapped opportunities of African markets. Corporate partnerships have mushroomed and some of technology's big players including Amazon, Google, Facebook and Alibaba continue to increase their presence across Africa's technology ecosystem landscape.
In its 2017 report The World Bank cited industry cloud and big data analytics as the next phase of the African digital revolution with the growth of IoT and the copious amounts of data to be generated on a daily basis, adding that the connectivity, analytical and cognitive services to be enabled will allow for the development and deployment of solutions to be achieved faster and more efficiently to boost growth, expand opportunities and improve service delivery.
In this context, it is expected that investments in highly scalable cloud computing to process and store this data; sophisticated big data analytics tools and platforms to extract the data value and insights; and security to protect the data, will steadily rise. For industry Cloud giants like Google Cloud, Amazon web Services, IBM Cloud, Oracle etc. looking to expand its big data and cloud services portfolio into the African continent, the biggest challenge is identifying the most ideal region to locate their business. With IT solutions no longer exclusive to the IT industry, locating industry cloud business in the heart of Africa's digital innovative hub is most essential to derive synergies from economies of scale, scope

and connectivity from the agglomeration and economic cluster of infrastructure, business institutions and technology ecosystem.

C. Interested Stakeholders

The work done in this project presents a significant opportunity for organizations in nearly every industry to leverage location data and spatial analytics as part of their everyday decision-making process to yield a new spectrum of competitive business insight. In particular, this project is of prime interest to IoT & Big Data Analytics and RapidScale Industry Cloud solutions providers like Amazon, Google, Microsoft, IBM, Oracle, etc. More generally, this project work is appealing to the concept of relatability. What has been researched in this study will be of interest to other researchers and institution interested in Africa's Technology Ecosystem, and that will add incrementally, to the patchwork of research on the said topic.

II. DATA

The biggest challenge I faced when working on this project was the lack of a comprehensive, validated and updated database on the technology ecosystem landscape of Africa, primarily as a result of the small base of literature and research on Technology Hubs in Africa. This study builds on a previous body of work by BongoHive, which maintains a current list of Technology Hubs in the African Tech Ecosystem landscape along with its colocation data on a crowdsourcing model, https://africahubs.crowdmap.com/reports/download.

A. Data Pre-processing

The last mapping exercise done by the World Bank in September 2015 listed 117 tech hubs in Africa. As of July 2016, a research by the GSMA Ecosystem Accelerator suggested there were 314 active tech hubs in Africa (a number aligned with Disrupt Africa's prediction in March 2016 article). In its 2018 report, the GSMA Ecosystem Accelerator recorded 442 active tech hubs across the Africa continent, suggesting that 128 new tech hubs have been established across Africa in two years' time.

Without this comprehensive dataset, data was downloaded and scrapped from multiple sources and combined into one table for the project. BongoHive's dataset on Africa's technology ecosystem maintained by a crowdsourcing model is downloaded and complemented with other data sources. To clean the data, unverified and inactive technology hubs were dropped from the project dataset. The project dataset is then formatted and restructured to an Excel file for ease of data wrangling and reading. The project dataset file is uploaded to my Github for reference - /resources/African-Hubs_Geocoordinates.xlsx

Fig. 2 Africa's active technology hubs dataset uploaded and transformed into a data-frame

Read and Transform the African Technology Hubs geodata file intp a pandas dataframe so that it is in a structured format

```
[9]: TechHubs_df = pd.read_excel('/resources/African-Hubs_Geocoordinates.xlsx', sheet_name=0, header=0)
     TechHubs_df.to_excel('/resources/African-Hubs_Geocoordinates.xlsx', encoding='utf8')

10]: TechHubs_df.shape

10]: (215, 8)

11]: TechHubs_df.head(5)
```

11]:

	M49 Code	ISO-alpha3 Code	Country or Region	Name of Technology Hub	Ciategory	City, State	Longitude	Latitude
0	24	AGO	Angola	KiandaHub	Business Incubator, Hacherspace/Makerspace, Co...	Luanda	13.234320	-8.836800
1	204	BEN	Benin	Centre Songhaï	Co-working	Porto-Novo	2.605000	6.497222
2	204	BEN	Benin	e-TRILABS	Business Incubator, Co-working	Cotonou	2.425500	6.362500
3	72	BWA	Botswana	Botswana Innovation Hub	Business Incubator, Co-working	Gaborone	25.917644	-24.679921
4	854	BFA	BurkinaFaso	Jokkolabs Ouagadougou	Business Incubator	Ouagadougou	-1.523372	12.354395

Explore and clean the African Technology Hubs datasats

```
12]: TechHubs_df.rename(columns={'Country or Region':'Country', 'Name of Technology Hub':'Hubs', 'Ciategory':'Category', 'City, St
     TechHubs_df.columns

12]: Index(['M49 Code', 'ISO-alpha3 Code', 'Country', 'Hubs', 'Category', 'City',
            'Longitude', 'Latitude'],
           dtype='object')

13]: print(type(TechHubs_df.columns))
     print(type(TechHubs_df.index))
```

B. Data Analysis

In this section, we explore and analyze Africa’s technology ecosystem landscape across slices of major dimensions to understand its evolution and growth, top ecosystem countries and cities.

Fig. 2.1 Digital Mapping of Africa’s technology ecosystem landscape

```
African_Hubs_map
```

[19]:

Africa’s Technology ecosystem Landscape Visualized

Fig. 2.2 Analyzing Africa’s Top 5 technology ecosystem Countries

```
print('Top 5 Ecosystem Countries by number of active hubs are:')
print('  South Africa', Top5_Countries.loc[0, 'Count'])
print('  Ghana', Top5_Countries.loc[1, 'Count'])
print('  Egypt', Top5_Countries.loc[2, 'Count'])
print('  Kenya', Top5_Countries.loc[3, 'Count'])
print('  Nigeria', Top5_Countries.loc[4, 'Count'])
print('Together, the Top5 Ecosystem by Countries account for over 50% of all active tech hubs in Africa.')
```

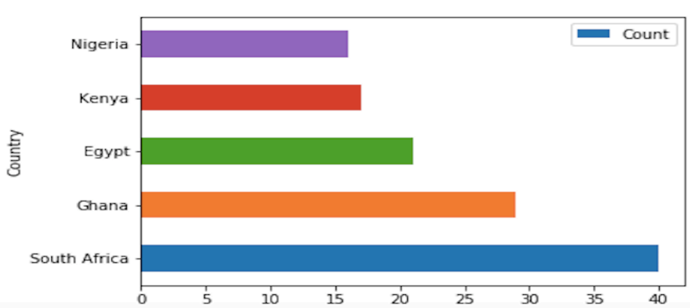
```
There are 33 Countries with Active Technology hubs in Africa:
Top 5 Ecosystem Countries by number of active hubs are:
  South Africa 40
  Ghana 29
  Egypt 21
  Kenya 17
  Nigeria 16
Together, the Top5 Ecosystem by Countries account for over 50% of all active tech hubs in Africa.
```

[27]:

```
Top5_Countries.plot(x='Country', y='Count', kind='barh')
```

[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f365062d8d0>
```



Africa’s Top 5 technology ecosystem countries account for over 50% of the active technology hubs

Fig. 2.3 Analyzing Africa's Top 5 technology ecosystem Cities

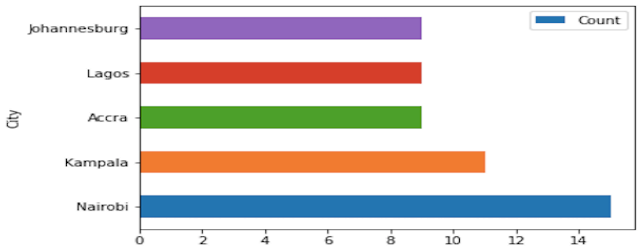
```
[32]: print('Africas {} Active Technology Hubs are spread between {} Cities'.format(CopyTechHubs_df['Hubs'].unique().shape[0],
           CopyTechHubs_df['City'].unique().shape[0]))

      print('The Top 5 Ecosystem Cities by number of active tech hubs are:')
      print('  Nairobi', Top5_Cities.loc[0, 'Count'])
      print('  Kampala', Top5_Cities.loc[1, 'Count'])
      print('  Accra', Top5_Cities.loc[2, 'Count'])
      print('  Lagos', Top5_Cities.loc[3, 'Count'])
      print('  Johannesburg', Top5_Cities.loc[4, 'Count'])
```

```
Africas 215 Active Technology Hubs are spread between 101 Cities
The Top 5 Ecosystem Cities by number of active tech hubs are:
   Nairobi 15
   Kampala 11
   Accra 9
   Lagos 9
   Johannesburg 9
```

```
[33]: Top5_Cities.plot(x='City', y='Count', kind='barh')
```

```
[33]: <matplotlib.axes._subplots.AxesSubplot at 0x7f364eeb2128>
```



The Top 5 African cities leading the continent's digital revolution

C. Data Exploration

The Foursquare API is leveraged for the exploratory data analysis of Africa's technology hubs ecosystem landscape dataset. For each unique technology hub, we identify and extract location/venues data on business interests and venues within a specified mile radius for insights and geographical implications. By the spatial analytics enabled by the Foursquare API, we are able to understand how Africa's digital ecosystem – start-up ecosystems, or networks of star-ups and their support structures – interact with each other, and with their immediate non-technical factors- cities and business institutions.
Following the location data/spatial analytics, the Foursquare API returned 224 unique categories of businesses and a total of 2304 business venues. The business venues location data was extracted from the Foursquare API and merged with the Africa's active technology hubs dataset into a data frame for further exploration.

Fig. 2.4 Analyzing the location data returned by the Foursquare API

Analyze Business Venues data

In order to get a better sense of the best way to generate Economic clusters, it's necessary to analyze the venues data returned by Foursqu

```
[69]: print('The total number of Businesses and Economic interests returned is ', Business_venues.shape[0])

      Business_venues.head(10)
```

```
The total number of Businesses and Economic interests returned is  2304
```

[69]:

	Hub	Hubs Latitude	Hubs Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	KiandaHub	-8.836800	13.23432	Pimms	-8.832830	13.236288	Mediterranean Restaurant
1	KiandaHub	-8.836800	13.23432	Cafe de Paris	-8.833305	13.236464	Coffee Shop
2	KiandaHub	-8.836800	13.23432	Restaurante NIHAO	-8.835894	13.232365	Asian Restaurant
3	KiandaHub	-8.836800	13.23432	Pastelaria Alvalade	-8.833648	13.235086	Bakery
4	KiandaHub	-8.836800	13.23432	Hotel Alvalade	-8.835543	13.233287	Hotel
5	KiandaHub	-8.836800	13.23432	BEL France	-8.833300	13.235574	Falafel Restaurant
6	KiandaHub	-8.836800	13.23432	Kero Gika	-8.835524	13.236366	Big Box Store
7	KiandaHub	-8.836800	13.23432	Hotel Fórum	-8.838022	13.234589	Hotel
8	KiandaHub	-8.836800	13.23432	Seven Restaurante	-8.833197	13.231855	Restaurant
9	Centre Songhaï	6.497222	2.60500	La Cabane De Sisko	6.496436	2.602253	Bar

```
[70]: print('Total Active Technoogy Hubs:', hubs_Centroids['Hub'].shape[0])
      print('Total Active Technology Hubs with Business venues:', Business_venues['Hub'].unique().shape[0])
```

```
Total Active Technoogy Hubs: 215
Total Active Technology Hubs with Business venues: 176
```

Location data with Venues, Venues Categories and Venues geo-coordinates returned by the Foursquare API

III. METHODOLOGY

In this project the k-means clustering algorithm is used to cluster Africa's technology ecosystem landscape.
The goal of this unsupervised machine learning technique is to segment and cluster Africa's business and technology ecosystem landscape by their unique attributes using the geographically referenced location data generated by the Foursquare API. By generating these distinct economic clusters, we are able to further explore our dataset for insights into their geographic and economic influences.

A. The k-means Clustering Algorithm

The k-means algorithm is an iterative algorithm that tries to partition a dataset into k pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data-points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data-points to a cluster such that the sum of the squared distance between the data-points and cluster's centroids (arithmetic mean of all the data-points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogenous (similar) the data-points are within the same cluster.

The k-means algorithm works as follows:

- Determine the number of clusters. Let it be k
- The initial step is to choose a set of k instances as centers of the clusters
- Next, the algorithm considers each instance and assigns it to the cluster which is closer
- The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment
- The process is iterated. K-means algorithm complexity O (tkn), where n is instances, c is clusters, and t is iterations. It often terminates at a local optimum.

The Elbow Method – Finding the value of k

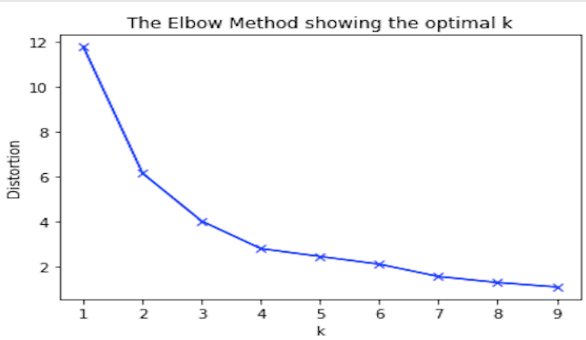
The Elbow method gives us an idea on what a good k number of clusters would be based on the sum of squared distance (SSE) between data-points and their assigned clusters' centroids. We pick k at the point where the SSE starts to flatten out and forming an elbow. Below is an illustration of the Elbow method using Africa's technology ecosystem landscape dataset.

Fig. 3.1 Finding the k value by the Elbow Method

```
from sklearn import metrics
from scipy.spatial.distance import cdist

distortions = []
K = range(1,10)
for k in K:
    k_means = KMeans(init = "k-means++", n_clusters = k, n_init = 12)
    k_means.fit(Business_Density['Business Density'].values.reshape(-1, 1))
    distortions.append(sum(np.min(cdist(Business_Density['Business Density'].values.reshape(-1, 1),
                                        k_means.cluster_centers_, 'euclidean'), axis=1)) / Business_Density['Business Dens

plt.plot(K, distortions, 'bx-')
plt.xlabel('k')
plt.ylabel('Distortion')
plt.title('The Elbow Method showing the optimal k')
plt.show()
```



Using the elbow method, the optimal value of the number of cluster was defined as 5 Run k-means to segment the Technology Hubs into 5 clusters.

The value of k identified as k = 5 from the Elbow method.

IV. RESULTS

The results from the exploratory analysis of Africa's technology ecosystem landscape by way of the unsupervised machine learning k-means clustering algorithm and spatial analytics/location data of the Foursquare API is presented below.

Fig. 4.1 Analyzing the venues and location data returned by the Foursquare API

```
[70]: print('Total Active Technoogy Hubs:', hubs_Centroids['Hub'].shape[0])
      print('Total Active Technology Hubs with Business venues:', Business_venues['Hub'].unique().shape[0])

      Total Active Technoogy Hubs: 215
      Total Active Technology Hubs with Business venues: 176

[71]: Hubs_diff = np.setdiff1d(hubs_Centroids['Hub'].values,Business_venues['Hub'].unique())
      hubs_Centroids[hubs_Centroids['Hub']==Hubs_diff[0]]
```

[71]:

	Areacode	ISO-alpha3 Code	Hub	City	Longitude	Latitude	Distance
137	686	SEN	Africa Living Lab Senegal	Egham Hill	-17.315463	14.773971	1643.0

```
[72]: Hubs_diff.shape

[72]: (39,)

[74]: Hubs_diff

[74]: array(['Africa Living Lab Senegal', 'Agro-Hub', 'Bauleni Social Centre',
             'Bloemfontein Central University of Technology Fab Lab',
             'Bright Youth Council Fab Lab', 'Center4Tech',
             'Centre for Entrepreneurship, Research & Innovation, Catholic University Institute of Buea',
             'Centre for Information Technology and Development (CITAD)',
             'Dare to Innovate', 'Digitising Cameroon Laws',
             'Etwatwa Intel Computer Clubhouse Network',
             'FABrication LABoratory Tanger (Gzenaya)', 'FabLab Addis',
             'GRATIS Central', 'GRATIS Eastern',
             'GRATIS Rural Technology Service Centre Ashanti Mampong',
             'GRATIS Rural Technology Service Centre Techiman', 'GRATIS Volta',
             'Great Kosa Suaye Technological Centre (workshop of Apostle Safo Kantanka)',
             'Innovation Lab', 'Jokko Labs Saint-Louis', 'Jokkolabs Abidjan',
             'KickAssTechnology-Obviously (KAT-O)',
             'Kimberly Northern Cape Higher Education Institute FabLab',
             'LakeHub', 'Limpopo Fablab', 'Ouaga Fablab', 'Ovillage',
             'Port Harcourt Hackerspace', 'SPARK',
             'Sedibeng Municipality FabLab',
             'SoftStart Technology Business Incubator',
```

Of the 215 active tech hubs in Africa's technology ecosystem, only 176 hubs had verified business venues within the specified radius

Fig. 4.2 Distribution of technology hubs by Clusters

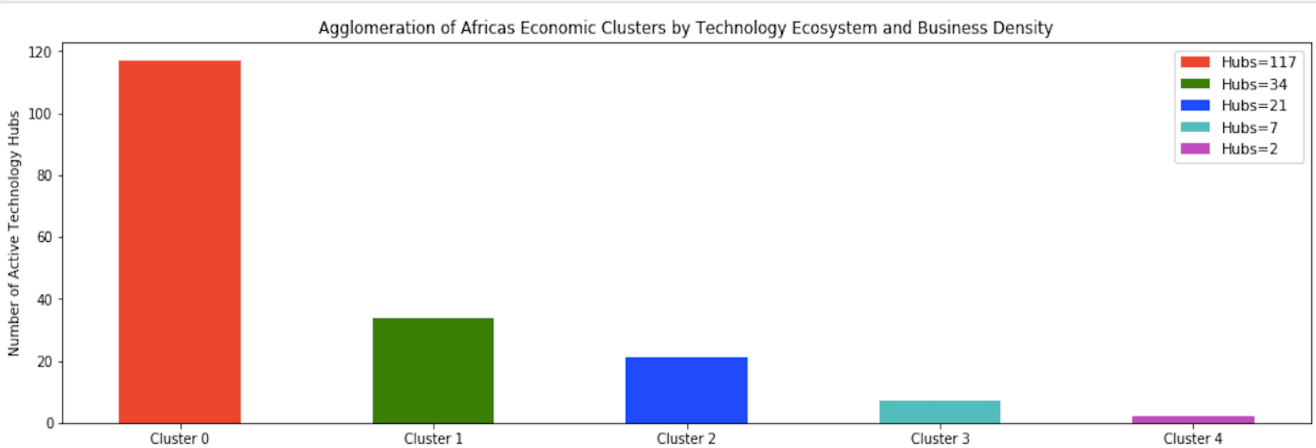
```
[95]: fig, ax = plt.subplots(figsize=(17,5))
      clusters= ['Cluster 0','Cluster 1','Cluster 2',
                 'Cluster 3','Cluster 4']
      colors = ['r','g','b','c','m']
      legend = []

      for id, cluster in enumerate(clusters):
          legend.append('Hubs={}'.format(cluster_info['Hubs'][id]))
          bar = ax.bar(cluster, cluster_info['Hubs'][id], 0.48, color=colors[id])

      ax.set_ylabel('Number of Active Technology Hubs')
      ax.set_title('Agglomeration of Africas Economic Clusters by Technology Ecosystem and Business Density')
      # ax.set_facecolor((1,1,1))
      # ax.grid(True, axis='y',color='k', linestyle='-', linewidth=1)

      plt.legend(legend, fontsize=11)

      plt.show()
```



Visualizing the distribution of Africa's technology hubs over the k = 5 economic clusters

Fig. 4.3 Implementation of the k-means Clustering Algorithm

Generate k = 5 economic clusters from the business and technology ecosystem dataset

```
[100]: # set number of clusters
kclusters = 5

Business_clustering = Business_Clusters.drop('Hub', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Business_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
[100]: array([0, 0, 4, 4, 4, 4, 4, 4, 4, 4], dtype=int32)
```

```
[101]: Hubs_Business_sorted.shape
```

```
[101]: (176, 11)
```

```
[102]: Top10_venues = Hubs_Business_sorted.copy()
```

```
[103]: Top10_venues['Cluster Labels'] = kmeans.labels_
Top10_venues.head()
```

[103]:

	Hub	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
0	302Labs	Café	Supermarket	Italian Restaurant	Toy / Game Store	Dessert Shop	Paper / Office Supplies Store	Squash Court	Juice Bar	Seafood Restaurant	Egyptian Restaurant	0
1	3abbar	Plaza	Café	Kebab Restaurant	Outdoor Sculpture	Cricket Ground	Dutch Restaurant	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	0

Africa's business and technology ecosystem dataset clustered into k = 5 clusters. Cluster labels: 0, 1, 2, 3, 4

V. DISCUSSION

Analysis of the project research results is presented in this section. Following the upload of the African technology ecosystem dataset, exploratory data analysis was performed across slices of major dimensions of Africa's technology ecosystem landscape to understand its evolution and growth. In Africa, although close to 50% of the active tech hubs are concentrated in 5 countries (South Africa, Kenya, Nigeria, Egypt and Ghana), almost each of the other African countries have at least one or two active tech hubs. Moreover, we have identified some leading countries by sub-region when it comes to tech hubs: Morocco and Egypt in North Africa; Nigeria, Ghana and Senegal in West Africa; Kenya and Uganda in East Africa; and South Africa in Southern Africa. Nairobi, Kampala, Accra, Lagos and Johannesburg were identified as Africa's top technology ecosystem cities leading the continents digital revolution.

Though very insightful, the results from this initial exploratory data analysis neglects an important aspect of business intelligence, the geographic and economic influences. Thus, the Foursquare API was leveraged for a more comprehensive exploratory data analysis of the project dataset. The Foursquare API applies geographical/spatial analytics and context to our project dataset for a thorough understanding of the influences of geography. Following this, we extract location data comprising 2304 business and economic interests of 224 unique venue categories, and the venues geo coordinates data.

The unsupervised machine learning technique, k means clustering is implemented to generate economic clusters from the projects business venues and technology hubs dataset by their unique discriminating features. These unique economic clusters are further examined and explored for a more comprehensive analysis and insight on their geographic and economic implications, as well as the influences of scope and connectivity.

A comprehensive analysis of each unique economic cluster is presented below. Each unique economic cluster is explored, classified and assigned a name based on its unique defining features and agglomeration of technology hubs and business venues.

Cluster 0 –

Classified as ***High-Density Digital Hub***

Cluster 0 has a total of 182 technology hubs, business and economic interests.

Fig. 5.1 Geographic and Economic Implications of Cluster 0

```
[148]: Cluster_0 = Economic_Cluster_df.loc[Economic_Cluster_df['Cluster Labels'] == 0, Economic_Cluster_df.columns[[0, 1, 2, 11, 12]]]
       Cluster_0
```

[148]:

	Hubs	# of Business venues	1st Most Common Venue	10th Most Common Venue	Cluster Labels
0	302Labs	16	Café	Egyptian Restaurant	0
1	3abbar	6	Plaza	Farmers Market	0
15	Alexandria Hackerspace	6	Plaza	Farmers Market	0
17	BIL Conference Tunisia	5	Café	Event Space	0
41	Dar Tekinohama Business Incubator (DTBI)	6	Café	Farmers Market	0
45	El-Minya Hacker space	4	Café	Falafel Restaurant	0
51	Fab Lab Egypt	29	Café	Bakery	0
54	FabLabTanger (ENSAT)	4	Plaza	Farmers Market	0
55	Fablab Tembisa (Ekurhuleni fablab)	1	Café	Event Space	0
71	HackZone	2	Café	Event Space	0
93	Karakeeb	6	Plaza	Farmers Market	0
124	SIDO - Dar es Salaam Regional Office	6	Café	Farmers Market	0
135	Tahrir Square	6	Plaza	Farmers Market	0
145	Thokoza Siyafunda Community Technology Centre ...	1	Café	Event Space	0
170	icecairo	69	Café	Plaza	0

```
[135]: print('Cluster 0: Geographic/Economic Implications:')
       print('    {}  Active technology Hubs'.format(Cluster_0['Hubs'].count()))
       print('    {}  Business venues and interests'.format(Cluster_0['# of Business venues'].sum(axis=0)))

       print('Cluster 0 classified as HIGH-DENSITY DIGITAL HUB based on its unique discriminating features and economic clusters')

       Cluster 0: Geographic/Economic Implications:
           15  Active technology Hubs
           167  Business venues and interests
       Cluster 0 classified as HIGH-DENSITY DIGITAL HUB based on its unique discriminating features and economic clusters
```

Cluster 1 –

Classified as a ***Low-Density Digital Hub***

Cluster 1 has 4 Active technology hubs and 10 Business and economic interests.

Fig. 5.2 Geographic and Economic Implications of Cluster 1

CLUSTER 1

```
136]: Cluster_1 = Economic_Cluster_df.loc[Economic_Cluster_df['Cluster Labels'] == 1, Econ
      Cluster_1
```

136]:

	Hubs	# of Business venues	1st Most Common Venue	10tl
47	Emerging Leaders in Technology and Engineering...	3	Bus Station	
61	GRATIS Northern	2	Record Shop	
64	GRATIS Upper West	1	Bus Station	
136	Takoradi Technical Institute Fab Lab	4	Bus Station	

```
138]: print('Cluster 1: Geographic/Economic Implications:')
      print('    {}  Active technology Hubs'.format(Cluster_1['Hubs'].count()))
      print('    {}  Business venues and interests'.format(Cluster_1['# of Business venues

      print('Cluster 1 classified as LOW-DENSITY DIGITAL HUB based on its unique discrimin

      Cluster 1: Geographic/Economic Implications:
          4  Active technology Hubs
          10  Business venues and interests
      Cluster 1 classified as LOW-DENSITY DIGITAL HUB based on its unique discriminating f
```

Cluster 2 –
Classified as ***Medium High-Density Digital Hub***
Cluster has 12 Active technology hubs and 33 Business venues and economic interests

Fig. 5.3 Geographic and Economic Implications of Cluster 2

CLUSTER 2

[139]:
```
Cluster_2 = Economic_Cluster_df.loc[Economic_Cluster_df['Cluster Labels'] == 2, Economic_Cluster_df.columns[[0, 1, 2, 11, 
Cluster_2
```

[139]:

	Hubs	# of Business venues	1st Most Common Venue	10th Most Common Venue	Cluster Labels
20	Bantalabs Senegal	2	Bar	Falafel Restaurant	2
29	Burundi Business Incubator (BBIN)	3	Hotel	Falafel Restaurant	2
32	CTIC DAKAR	2	Hotel	Falafel Restaurant	2
35	Centre Songhaï	1	Bar	Falafel Restaurant	2
49	FABlab Namibia	4	Restaurant	Falafel Restaurant	2
104	Mobile Web Ghana	7	Theater	Flea Market	2
105	MobileSenegal	2	Hotel	Falafel Restaurant	2
110	Namibia Business Innovation Centre (NBIC)	4	Hotel	Event Space	2
112	North West University	2	Bar	Event Space	2
125	Sahara Labs - Tarfaya,Hackerspace	1	Hotel	Event Space	2
155	West African Agribusiness Resource Incubator (...	2	Hotel	Event Space	2
163	e-TRILABS	3	African Restaurant	Falafel Restaurant	2

[141]:
```
print('Cluster 2: Geographic/Economic Implications:')
print('    {}  Active technology Hubs'.format(Cluster_2['Hubs'].count()))
print('    {}  Business venues and interests'.format(Cluster_2['# of Business venues'].sum(axis=0)))

print('Cluster 2 classified as MEDIUM HIGH-DENSITY DIGITAL HUB based on its unique discriminating features and economic clu
```

```
Cluster 2: Geographic/Economic Implications:
    12  Active technology Hubs
    33  Business venues and interests
Cluster 2 classified as MEDIUM HIGH-DENSITY DIGITAL HUB based on its unique discriminating features and economic clusters
```

Cluster 3 –
Classified as Very Low-Density Digital Hub
Cluster 3 has 3 Active technology hubs and 7 Business and economic interests.

Fig. 5.4 Geographic and Economic implications of Cluster 3

CLUSTER 3

[142]:
```
Cluster_3 = Economic_Cluster_df.loc[Economic_Cluster_df['Cluster Labels'] == 3,
Cluster_3
```

[142]:

	Hubs	# of Business venues	1st Most Common Venue	10th Most Common Ven
58	G Space, Busy Internet	2	Cruise	Farmers Mar
77	Hubsocial	3	Cruise	Falafel Restaur
160	Zamrize	2	Cruise	Farmers Mar

[143]:
```
print('Cluster 3: Geographic/Economic Implications:')
print('    {}  Active technology Hubs'.format(Cluster_3['Hubs'].count()))
print('    {}  Business venues and interests'.format(Cluster_3['# of Business v

print('Cluster 3 classified as VERY LOW-DENSITY DIGITAL HUB based on its unique
```

```
Cluster 3: Geographic/Economic Implications:
    3  Active technology Hubs
    7  Business venues and interests
Cluster 3 classified as VERY LOW-DENSITY DIGITAL HUB based on its unique discri
```

Cluster 4 –
Classified as ***Very High-Density Digital Hub***
Cluster 4 is Africa's highest density of business venues and economic interests (2087) and technology hubs (142). This is Africa's prime digital innovation hub leading the continents digital revolution with an unfulfilled promise of integrating the continent into one large network to enable abundance and spur economic growth. This is **Silicon Sahara**.

Fig. 5.5 Geographic and Economic implications of Cluster 4

CLUSTER 4

```
[147]: print(' {} Active technology hubs in Cluster 4'.format(Cluster_4.shape[0]))

       Cluster_4.head()

 142 Active technology hubs in Cluster 4
```

[147]:

	Hubs	# of Business venues	1st Most Common Venue	10th Most Common Venue	Cluster Labels
2	88mph Garage, Nairobi	7	Golf Course	Electronics Store	4
3	AKENDEWA TECH HUB	2	Clothing Store	Farmers Market	4
4	AMN Co Working Space	1	Boat or Ferry	Falafel Restaurant	4
5	ARO Fablab	5	Soup Place	Eastern European Restaurant	4
6	AUC Venture Labs	68	Café	Fast Food Restaurant	4

```
[145]: print('Cluster 4: Geographic/Economic Implications:')
       print('    {}  Active technology Hubs'.format(Cluster_4['Hubs'].count()))
       print('    {}  Business venues and interests'.format(Cluster_4['# of Business venues'].sum(axis=0))

       print('Cluster 4 classified as VERY HIGH-DENSITY DIGITAL HUB based on its unique discriminating fea

Cluster 4: Geographic/Economic Implications:
    142  Active technology Hubs
    2087  Business venues and interests
Cluster 4 classified as VERY HIGH-DENSITY DIGITAL HUB based on its unique discriminating features a
```

VI. CONCLUSION

In this project work is done exploring the technology hubs of Africa's emerging digital ecosystem based on actual economic and geographic implications, by leveraging spatial analytics and location data enabled by the Foursquare API and an unsupervised machine learning classification methodology, the k-means clustering algorithm.
The project was very conclusive in affirming the applicability of the k-means clustering algorithm and location data/spatial analytics to maximize the value of business intelligence by revealing insights that drive better decision making.
This report outlines initial finding across a subset of active technology hubs in Africa, and it prove that the research methodology was very effective in generating economic clusters featuring high homogeneity of the segments, and high heterogeneity among segments. Integrating location data and spatial analytics of the Foursquare API was very critical in revealing the relationships between technology hubs, start-up ecosystems and their support structures, and their cities, infrastructure and business institutions. Following this comprehensive analysis, Africa's prime digital innovation hub – **Silicon Sahara**, is identified as predicated on geographic and spatial factors. Interactive maps are used to present research results to expose spatial relationships and influences that just aren't visible in traditional tabular views of data, thereby allowing stakeholders to visualize the influence of geography on the technology ecosystem and economic clusters.
Some challenges that have affected the depth of the findings of this report include a lack of validated and updated dataset on the technology ecosystem landscape of Africa. This report outlines initial findings across a subset of Africa's technology landscape dataset. The conclusions are based on an extensive review of this subset dataset, which means conclusions only apply within these

limitations of data outsourcing constraints. This must be considered in the evaluation of the presented findings and the planning of future work. Future studies will examine this topic with a more complete dataset.

VII. APPLICATION

Every business event or transaction happens at a specific location for reasons that, when clearly understood, reveal insights that comprise the "intelligence" component of Business Analytics. There are location and geographic context to every element of business. Ignoring this leads to substandard decision making by businesses and stakeholders.

The work done in this project presents a significant opportunity for organizations in nearly every industry to leverage location data and spatial analytics as part of their everyday decision-making process to apply geographical/spatial context to information to inform actions or responses to business opportunities, thereby yielding a new spectrum of competitive business insight.

Business intelligence is all about extracting high-value information from vast amounts of data and presenting it to the right people in a highly consumable format. Today, forward-looking organizations are learning firsthand how adding spatial analytics and interactive mapping to their BI systems can drive powerful synergies for a more comprehensive and insightful data-driven decision making. Having a unified view of information that includes the locations of your products and their places in the supply chain can enable you to monitor the continuity of business and proactively manage its efficiency through delays or crises.