BIG DATA
CASE Study

WORLD BANK GROUP

Using Big Data to Measure City Innovation Capacity

Megha Mukim and **Juni Zhu**

SUMMARY

City leaders around the world have been grappling with economic development challenges in the face of slowing growth, changing demographics, and increasing unemployment rates, especially among youth. Task teams at the WBG are searching for ways to better understand these challenges and find solutions to help their clients.

Private-sector firms are the main drivers of job creation, productivity, and wage increases; they also drive much innovation. Despite the role that privately-held start-ups play in innovation-led growth, cities currently lack reliable, up-to-date, and comparable data necessary to understand and inform policy decisions that affect startups. The Start-Up City Dashboard aims to (1) provide reliable, comprehensive, and comparable data on start-up activity and innovation ecosystems in data scarce environments; (2) provide a better understanding of start-up activity drivers to guide more targeted policy; and (3) demonstrate the use of big-data tools for more standardized WBG data and analysis.

The Start-Up City Dashboard is comprised of three interactive visual diagnostic tools that help measure and compare the health, diversity, and scale of innovative economic activity in 22 pilot cities: the Health of the Innovation Ecosystem Tool, Industry Benchmarking and Uniqueness Dashboard, and the Innovation Archetype City-by-City Comparison. These tools allow city governments to obtain an up-to-date and accurate picture of their innovation ecosystems and to learn from other cities that are operating differently.

CHALLENGE

The Trade & Competitiveness Global Practice's report, "Competitive Cities for Jobs and Growth: What, Who and How," aims to help cities understand how to facilitate private sector growth to create jobs, raise productivity, and increase incomes. The report identified four enabling factors for growth: (1) institutions and regulations; (2) infrastructure and land; (3) enterprise support and finance; and (4) skills and innovation. The findings also suggested that the creation of innovative small firms and the displacement of incumbents was one of the main sources of innovation and – according to the team's experience – a topic of interest for many city governments. City leaders asked three key questions related to the role start-ups play in innovation-led growth: "Who are the entrepreneurs and start-ups in my city?"; "What industries do they focus on, and are these unique to my city?"; and, "How is my city doing compared to others?"

Cities lack reliable, up-to-date, and comparable data on their innovation ecosystems that would help them answer these questions. Moreover, readily-available data sources are not much help for a variety of reasons. For example, most available data is often aggregated at the national level. Even when sub-national data is available, it is often limited to industrial sectors. It is difficult to find data necessary to assess the factors that contribute to a successful start-up ecosystem in cities, especially those that are intangible in nature, such as networking assets to help entrepreneurs get connected, or a city culture that tolerates failures and encourages collaboration.

INNOVATION

To provide cities with a reliable diagnostic tool, the team gathered data on each of the following factors that contribute to a successful start-up ecosystem: human capital, financial infrastructure, urban amenities, collaborative culture, and networking assets. They did this by identifying proxy indicators for each of the factors for which open-source data could be updated frequently and rapidly, and by using a combination of data science tools including R, Python, STATA, Excel and Tableau for data collection, transformation, analysis, and visualization. For example:

- Strength of human capital in a city is determined by the number of universities as obtained in real-time from Open Street Maps.

- Financial infrastructure is captured by the number of banks or ATMs in a city as obtained in real-time from Open Street Maps.

- The nature of networking assets can be assessed through information on available networking activities, for example, incidence of mentee-mentor relationships and presence of serial investors. The dashboard used AngelList to determine the percentage of entrepreneurs who are well-connected as a proxy for the strength of a city's networking assets.

- Urban amenities is approximated by the ubiquity of coffee-shops, pubs and restaurants as obtained in realtime from Open Street Maps.

- Lastly, collaborative culture is assessed by examining the percentage of technicians who are active on collaborative online platforms, such as Stack Overflow.

By virtue of being problem- and demand-driven, the project benefited from unique approaches in both design and process. Rather than being a purely academic endeavor, the project had a concrete goal of addressing concerns that clients consistently brought up in consultations, such as whether start-ups in their cities are creating competitive jobs for young people. In addition to speaking with World Bank clients to inform the design process, the team also consulted other Global Practices and Regions within the WBG. These consultations lent relevance to the project and led to the enthusiastic reception from clients.

RESULTS

Three tools comprise the Start-Up City Dashboard, whose prototype has data on twenty-two pilot cities ranging from Dar es Salaam to New York:

1. The Global Start-Up City Snapshot provides a snapshot of the innovation ecosystem's health, including an overall rank and a breakdown score for each of the five factors that contribute to a start-up ecosystem.

2. The Industry Benchmarking and Uniqueness Tool allows for the identification of the industrial mix of the city's start-ups and how these industries compare with two to three similar cities.

3. The Archetypes of Innovation Activities Tool allows for one-to-one comparison of cities against the four innovation archetypes identified by the consulting firm, McKinsey & Company: science-based, engineering based, customer-focused, and efficiency-based. This tool indicates a city whether it is strong or weak in a particular type of innovative activity compared to competitor cities.

Sub-national clients have responded enthusiastically to the prototype dashboard. For example, city leaders in Shanghai were able for the first time to compare Shanghai's performance to Seoul, Tokyo, New York, and even other Chinese cities, disaggregate the data by sector, and ask what other cities might be doing differently. The Dashboard is now being piloted in Tanzania as part of a broader initiative to understand entrepreneurial ecosystems, leading to the design and preparation of a US$100 million lending operation.

The team is keen to build upon its success. First, by working with a capstone group, the team is focusing on understanding the direction and the magnitude of the possible bias using these new sources of data obtained through web scraping methods as compared to data obtained from other traditional sources in select OECD and upper middle-income cities. Second, the team is exploring opportunities for corporate partnerships with IBM and LinkedIn to further the work on gathering reliable data on entrepreneurship and to deepen and scale the Dashboard to include additional variables. Third, the team is looking to scale the project to include up to 600 cities worldwide, including many in low-income countries.

This project also illustrates that similar web scraping methods from open-source websites to obtain national or sub-national proxy data can be employed to develop monitoring and diagnostic tools for other projects within the Trade & Competitiveness Global Practice and the WBG. Once these tools are established, efforts to maintain and scale them could be marginal.

LESSONS LEARNED

1. ***Design based on a solid analytical foundation:*** This project was built on two years of initial research to understand the importance of helping clients with economic development challenges. The project brought together team members with skill sets in urban and private sector development, which was critical to developing a broad-based tool for task teams operating across different thematic areas.

2. ***Demand as a foundation for design:*** This project proved the importance of aligning design in response to demand. The initial thinking underlying the tool was based on increasing demands from clients to understand start-up activity, particularly from a project in China. Strong demand kept the project focused and ensured its outputs were impactful for and responsive to its end-users (i.e. city leaders). In addition, the team continually reached out to regional task teams and experts in other Global Practices (GP) (Information and Communications Technology (ICT) & Transport GP, Social, Urban, Rural, and Resilience GP) to solicit feedback, which made the tool flexible to the needs of different clients and users.

3. ***Seek help on technical expertise:*** This project was fairly new for the team and the WBG. As a result, the team faced many challenges, including finding the right technical skills in the absence of a standardized Terms of Reference for the required expertise. For this, the team turned to other advisors both within the Trade & Competitiveness Practice and the Big Data team, and even to private sector firms, for advice. It helped to have a clearly identified knowledge lead in the GP to provide guidance and regular feedback and support.

RESOURCES

Start-Up City Index, Health of the Innovation Ecosystem Tool:
https://public.tableau.com/profile/romulo.cabeza#!/vizhome/DashboardDraft/WholeDashboard

Industry Benchmarking and Uniqueness Dashboard:
https://public.tableau.com/profile/romulo.cabeza#!/vizhome/UniqueMarketsDashboardFinalversion/UniqueMarketsInformation

Innovation Archetype City-by-City Comparison:
https://public.tableau.com/profile/romulo.cabeza#!/vizhome/InnovationArchetypeCity-by-CityComparison/SequentialPresentation

Competitive Cities for Jobs and Growth: What, Who, and How:
http://documents.worldbank.org/curated/en/902411467990995484/pdf/101546-REVISED-Competitive-Cities-for-Jobs-and-Growth.pdf