

# Yelp Analysis

Zhixuan Chi, Hengzhi Wu, Nguyen Cao, Madana Krishnan V.K

CMPT732 Big Data Lab 1 - Course Project

December 3, 2019

# Overview

1 Problem

2 Datasets

3 System Architecture

4 Implementation

5 Results

Estimating GDP is an enormous task-doing so at county levels is even tougher:

- Due to data challenges. This means that GCP is the best available measure of aggregate activity at the county but not a perfect measure;
- Prevalence of informalities. Many transactions or activities in the economy, including many that go through the market are never recorded officially; and
- Difficulties in attributing economic activity to a specific county unit. It gets worse when we start to account for cross-county transactions.

There is great interest at counties to know the relative size of their economies.

Does the Mt Kenya Region contribute 60% of Kenya's GDP?

## What are the potential applications

There are numerous applications of this novel data set, including:

- GCP is an acceptable proxy for county's tax base—that could be leveraged for Own Source Revenue (OSR) mobilization;
- GCP is a great indicator for monitoring progress (changes in growth over-time) across counties and would potentially corroborate other HDI;
- The national government could use GCP to quantify the dividends from fiscal reforms; and
- Finally, GCP could be used to identify counties that require more focus for inclusive growth and pro-poor policies.

## There is significant differences in the size of the economy across counties

- Urban counties-Nairobi(21.7%); Nakuru(6.1%), and Mombasa (4.7%) vs. Isiolo(0.2%), Samburu(0.3%), and Lamu(0.4%);
- Although some counties have a small contribution to the national cake, they show great potential for catch-up. e.g. E.Makwet expanded at 10% per year.
- Agriculture remains a key driver of growth in most counties-thus initiatives to climate proof this sector could contribute greatly to poverty reduction; and
- Only about 25% of the counties have their GCP per capita above the national GDP per capita-showing great disparities but also what devolution brings on the table to tackle this.

## What are the user be aware warnings?

- Non-market production-It is difficult to measure or estimate the value of goods and services that are not sold in a market.
- Unrecorded activities both illegal and legal.
- Frequent revisions-preliminary estimates should always be used with caution.
- GCP is not a good measure of well being due to lack of accounting for externalities.
- GCP tells nothing about distribution of income within the county.

## My Takeaways: What gets measured gets done

- Good quality data not only measures progress, but it also inspires it.
- The publication of the GCP data has plugged an important data gap to track progress of devolution in Kenya.

The End