<u>**Readme**</u>

This note explains how to replicate the results in "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City".

The data and code required to replicate the estimates in the paper are contained within this folder.

The contents of the folders in this root directory are as follows.

- **data**: We provide cleaned individual data with all variables required to produce the results in the paper, but with remaining variables and contact details removed.
- **do**: We provide code to produce the results in the paper
- **utilities**: Here we provide additional code used to create the tables in the paper. In particular, we use specific programs to output complete tables to latex based on a single line of input code. In addition, we have provided additional ado files, and do files
- **figures:** Output figures are saved here
- **tables:** Output tables are saved here

Opening and running the code

All analysis is performed in Stata. We have saved all datasets in a format that can be read by Stata 11, 12 and all later versions. The code is contained in the folder "do". Each do file is self-contained. That is, it can be run independently and will output results into folders in the home directory. To run any do files, users need only to set their root directory to the root directory of this folder. For example, do files will call the data with the command: "use data/endline_data.dta" without any further root information. Since the do files call on various auxillary datasets in subfolders, it is important to set the root directory correctly.

We note that a number of publically available additional packages are used throughout our code and may need to be installed: est2vec, parmby, etc. To the best of our knowledge, all packages should be publically available and easy to install. One package we believe may be harder to find online was the *dataout* package, which is provided in the "utilities" folder.

Datasets

We have provided one main dataset that can be used to produce the main experimental results in the paper. "endline_data.dta" There are also a number of additional datasets required to perform additional analysis. These datasets are stored separately in order to keep the main data file streamlined.

Datasets can be merged on the unique identifier "r_id". This is an anonymized version of the personal identifier used in the study, and can be used to merge the data.

Summary of the main do files and the corresponding tables produced

file	tables produced
main_endline_results.do	Most of the main figures and tables in the paper that require only the two endline datasets merged with the baseline data. Code for all outputs should be found here unless otherwise specified below.
complete_family_tables_a15_a23.do	This file produces a comprehensive set of estimates on secondary outcomes pre-specified in our pre-analysis plan. These are contained in the Appendix, Tables A.15 to A.23

trajectories.do	This code produces all of the plots showing the trajectory of impacts over time using the full set of high-frequency phone survey data, stored in phone_panel.dta
mediation_analysis.do	This file produces figure 4 in the main paper. It needs to be run separately because it rearranges the two waves of data to merge the first endline with the second, at the individual level.
premia.do	This produces Table A.26 in the appendix. Like the mediation analysis do file above, this code needs to merge the two endlines together to regress wages in the second endline on employment outcomes at the first endline.
bounds.do	This file produces attrition_bounds.dta which includes individuals who were included in the baseline but not interviewed in follow-up surveys in order to (i) produce analysis of the determinants of attrition, and (ii) bounds for the magnitude of treatment effects under different assumptions about the patterns of attrition (found in Section A.2 of the online appendix)
spillovers.do	This file works with our sample of untreated individuals in treated areas, in order to estimate the local spillover effects of the interventions, as reported and presented in Section A.3 of the paper.
rdd_a14.do	This code creates Table A.14. This runs separately because this file draws on test dating from the testing component of the Workshop treatment.
endogenous_stratification.do	This creates the results contained in the final rows of Tables 4, 5, 6. The results are saved in output table table_4_5_6_finalrows.tex. We have then manually added those figures to the final rows of the respective tables. Note that because this exercise involves a bootstrap-like approach, the results are not deterministic, and may differ from instance to instance.
comparison_other_studies.do	This produces Figure 5. This draws on the file itt_other_studies.dta which contains ITT estimates from a number of recent studies in the literature. See the notes to Table 5 for more information on the criteria used to select these studies.