

Kiva loans: crowdfunding for good

Merola Paolo¹, Fratello Mattia², Bowen Sebastiano³

Abstract: Grazie alla Quarta Rivoluzione ed all'enorme connettività globale, il microprestito, inteso come prestito di bassa entità senza intermediazione di istituti di credito, è diventato un fenomeno distribuito globalmente, soprattutto con l'avvento di Internet. Nato come strumento per combattere la povertà nei paesi del Terzo Mondo, è stato al centro di numerosi studi volti a determinarne la reale efficacia, assumendo a volte connotati di puro investimento economico. Il microprestito è erogato senza alcun collaterale a garanzia del pagamento, quindi spesso il prestatore si assume un forte rischio di perdita della somma prestata; di conseguenza, il tasso di interesse può diventare molto elevato (il sito Prosper.com offre un interesse fino al 31.9% per i prestiti a più alto rischio) oppure il prestito assume connotati di beneficenza.

Nasce così nel 2005 Kiva, un'associazione nata nel 2005 per effettuare microloaning con fini benefici, secondo la formula del crowdfunding: quando una richiesta di prestito viene effettuata a Kiva o ad uno dei suoi Partner, essa viene proposta alla comunità di Lender (Prestatori/Creditori) e, se il prestito viene finanziato, viene poi erogato.

In questo progetto ci siamo proposti di creare un database dal quale analizzare i dati sui prestiti erogati tramite Kiva, in particolare relativi ai paesi più poveri del mondo. Abbiamo creato un database di documenti BSON, contenenti informazioni come la località, l'ammontare del prestito ed anche spiegazioni sulle motivazioni che hanno spinto il richiedente a rivolgersi a Kiva. Abbiamo inoltre arricchito i dati di Kiva, aggiungendo alcuni indicatori economici della Banca Mondiale, al fine di effettuare uno studio preliminare sull'efficacia di tale attività di microloaning nella lotta alla povertà, o comunque sull'influenza che questo flusso di denaro ha sui paesi riceventi.

¹Data Science, 834098 ²Data Science, 785142 ³Data Science, 718327

CONTENTS

I	Introduzione: il Microcredito	1
II	Scelta delle fonti, struttura dei dati e scelta del DB	1
III	Architettura del DB	2
IV	Flusso e struttura dei dati	3
V	Manipolazione dei dati	3
VI	Queries effettuate e risultati	4
VI-A	Query 1	4
VI-B	Query 2	4
VI-C	Query 3	5
	References	7
	Appendix	8

I. INTRODUZIONE: IL MICROCREDITO

Il fenomeno del microprestito inteso come prestito informale tra individui esiste da millenni, ma il microprestito "moderno" pare sia nato intorno agli anni '70 in Bangladesh. Durante una carestia il Dottore Muhammad Yunus, Professore di Economia all'Università di Chittagong, stava visitando vari villaggi locali nell'ambito di uno studio personale volto a capire le cause della diffusa povertà del suo paese. In uno di questi villaggi incontrò un gruppo di 42 donne che costruivano sgabelli di bamboo. Le donne, essendo prive di mezzi, erano costrette a chiedere i materiali in prestito ai commercianti, i quali esigevano in cambio l'esclusiva sull'acquisto degli sgabelli. Siccome, detratto il prestito dei materiali, i commercianti compravano gli sgabelli ad un prezzo bassissimo, si formava così un circolo vizioso che, all'atto pratico, riduceva le donne in condizione di semi-schiavitù.

A stupire il dottore fu l'ammontare del prestito che teneva le donne vincolate ai commercianti: 27\$. Il dottore fornì la somma di tasca sua dietro promessa di restituzione ad interesse zero, permettendo alle donne di rompere l'accordo e vendere i loro prodotti altrove a prezzo maggiore, uscendo dal circolo vizioso. Era nato il microloaning.

Successivamente il dottor Yunus fondò la Grameen Bank ¹, che oggi opera in oltre 80.000 villaggi e che, nel 2006, ha ricevuto il nobel per la pace assieme al suo creatore.

Nel 2005, quindi già pienamente nell'era dell'iperconnessione, Kiva inizia il servizio di microloaning, tramite il Web, seguendo la formula del crowdfunding. Quando un prestito viene richiesto ad un ente locale partner di Kiva, oppure a Kiva stessa senza intermediari, la richiesta viene sottoposta al vaglio della community di iscritti al sito di Kiva, i quali, sulla base di una descrizione dei motivi e degli scopi per i quali si richiede il finanziamento, avranno un certo periodo di tempo a partire dalla richiesta per decidere se contribuire (di norma con 25\$) a finanziare il prestito. Trattandosi di un ente benefico, Kiva non richiede interessi ed impone ai suoi partners locali di mantenere limitati i propri tassi, mentre Kiva stessa copre i propri costi operativi mediante donazioni dai suoi lenders o da fondazioni.

¹letteralmente "Banca del Villaggio"

II. SCELTA DELLE FONTI, STRUTTURA DEI DATI E SCELTA DEL DB

Il sito internet di Kiva espone delle API ad uso pubblico, come parte della loro politica di totale trasparenza, e forniscono anche degli snapshot periodici dei propri dati. Allo scopo di organizzare liberamente il nostro database, abbiamo preferito scaricare i dati grezzi mediante le API.

La seconda fonte di dati è la Banca Mondiale, che espone API ad uso pubblico dalle quali abbiamo scaricato i valori di 4 indicatori economici che abbiamo ritenuto rilevanti per la nostra analisi finale.

La fonte principale di dati è sicuramente Kiva, che fornisce un certo numero di dettagli sui micro-prestiti effettuati. Ogni documento esposto consiste in un singolo prestito, descritto mediante i campi nella tabella sottostante:

Campo	Descrizione del campo
Id	ID univoco del documento
funded amount	L'ammontare raccolto da Kiva per finanziare il prestito
location	Il paese dal quale arriva la richiesta di prestito
funded date	Data di finanziamento del prestito (assente se status = "fundraising")
name	Nome di riferimento del richiedente (persona o associazione)
status	Lo stato del prestito: "funded" oppure "fundraising"
image	ID e Template di un'immagine associata alla richiesta (non inclusa nei dati)
activity	Parola chiave sull'attività condotta dal richiedente
sector	Parola chiave sul settore relativo all'attività condotta dal richiedente
themes	Classificazione di Kiva dei settori
use	Lo scopo per il quale il richiedente intende usare il prestito
geo	Informazioni più dettagliate sulla località
posted date	Data nella quale la richiesta di prestito è stata effettuata
planned expiration date	Data prevista di maturazione del prestito
loan amount	L'ammontare richiesto in prestito
borrower count	Il numero di persone facenti parte della richiesta di prestito
lender count	Il numero di utenti Kiva che hanno finanziato il prestito
bonus credit eligibility	Un indicatore di affidabilità del richiedente
tags	Eventuali tags applicabili alla richiesta

TABLE I
BREVE DESCRIZIONE DEL SIGNIFICATO DEI CAMPI CONTENUTI NEI
DOCUMENTI SCARICATI

La fonte successiva di dati è la Banca Mondiale. Purtroppo, un'istituzione così grande ed importante espone i propri dati in modo altamente caotico. In particolare, le API permettono di scegliere tra 15874 indicatori, ognuno con un suo codice univoco. Tuttavia, di questi indicatori solamente poche centinaia sono effettivamente aggiornati ed utilizzabili, mentre moltissimi sono ripetuti oppure non mantenuti; il risultato è che richieste effettuate su questi ultimi restituiscono risultati inutilizzabili.

Ai fini della nostra analisi, abbiamo individuato 4 indicatori economici di interesse: GDP, GNI, GNI (metodo Atlas) e GNI pro capite.

- **GDP:**
Gross Domestic Product (PIL, prodotto Interno Lordo), il classico indicatore di prosperità economica di un paese definito come il valore di mercato aggregato di tutte le merci finite e di tutti i servizi prodotti nei confini di una nazione in un dato periodo di tempo. Pur essendo la base di ogni analisi economica su un paese, non è sufficiente ai fini di questo lavoro, motivo per il quale sono stati aggiunti altri indicatori;
- **GNI:**
Gross National Income (RNL, Reddito Nazionale Lordo), è un indicatore creato aggiungendo al GDP i movimenti di denaro tra il paese in oggetto ed il resto del mondo. Viene utilizzato massicciamente per decidere l'ammontare dei contributi al budget dell'Unione Europea da parte degli Stati Membri. In particolare, sottraendo il GDP dal GNI, si ottiene l'ammontare del flusso di denaro da e verso il paese in oggetto, quindi è risultato interessante confrontare questo indicatore con l'ammontare dei loans di Kiva;
- **GNI - metodo Atlas:**
Gross National Income - metodo Atlas, è lo stesso indicatore di cui sopra ma convertito in USD mediante il fattore di conversione Atlas ²;
- **GNI pro capite:**
il GNI di un paese diviso per il numero di abitanti nominalmente in età lavorativa.

Per ciascuno di tali indicatori abbiamo compilato una tabella con i valori per paese relativi agli anni dal 2005, anno di fondazione e di inizio attività di Kiva, al 2016. Purtroppo la Banca Mondiale non fornisce (o forse non possiede ancora) i dati relativi al 2017, quindi la comparazione dei dati dalle due fonti si è dovuta limitare a questi 12 anni.

Il database utilizzato per lo storage dei dati è MongoDB. Si tratta di un DB NOSQL document-based, nel quale ogni singolo elemento è salvato come documento BSON, quindi come lista di

coppie chiave-valore eventualmente annidate. Abbiamo preferito questo tipo di database in quanto gli attributi dei nostri documenti, seppure per la maggior parte combacino, sono a volte differenti. Ad esempio, il campo "funded date" contiene la data nella quale il prestito è stato finanziato, ed ovviamente non sarà presente in documenti che rappresentano prestiti non finanziati. Ci serviva quindi una struttura elastica in modo da evitare una tabella SQL piena di Null. Altro aspetto che abbiamo considerato nella scelta è il grande supporto e relativa semplicità d'uso di cui gode MongoDB rispetto ad altri sistemi come Cassandra o Neo4j; ad esempio, MongoDB supporta nativamente lo sharding, rendendo facile ed efficiente eliminare i single point of failure che affliggono i database distribuiti.

III. ARCHITETTURA DEL DB

Il database è stato costruito per garantire scalabilità e riproducibilità, secondo gli standard di un DB distribuito.

- **Scalabilità:**
invece di una singola istanza di mongod, il db è stato frammentato in 3 pezzi, detti shards, collegati tutti ad una quarta istanza che funge da configuration server (ovvero il db che conosce in quale shard si trovi ogni dato). Il configuration server e tutti gli shards sono stati poi collegati ad un router mongos, che opera sugli shards in base alle richieste esterne sfruttando le informazioni contenute nel configuration server. Il router mongos è quindi l'unico contatto tra il DB e gli utenti. Da notare che il configuration server ed il router sono entrambi single point of failure, quindi di norma se ne implementano più di uno. In questo progetto ciò non è stato fatto, lasciando alla riproducibilità il compito di garantire la conservazione dei dati e l'accessibilità.
- **Riproducibilità:**
ogni shard ed il config server non sono singole istanze ma insieme di 3 copie dello stesso database; tale insieme è detto replica set. MongoDB marca una delle copie come primary e le altre come secondary; sia la scrittura che la lettura di dati avverranno di norma sulla copia

²<https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-the-world-bank-atlas-method-detailed-methodology>

primary, passando alle copie secondary solo in caso la primary non sia disponibile. A scopo di test e dimostrazione, tutti gli shards e le repliche sono state avviate sulla stessa macchina: in fase di utilizzo reale, i vari shards e repliche sarebbero, appunto, distribuiti su varie macchine: l'ideale sarebbe avere 3 macchine collegate in modo che ciascuna abbia una delle copie di ciascun replica set + un config server + un router mongos, così da eliminare ogni point of failure.

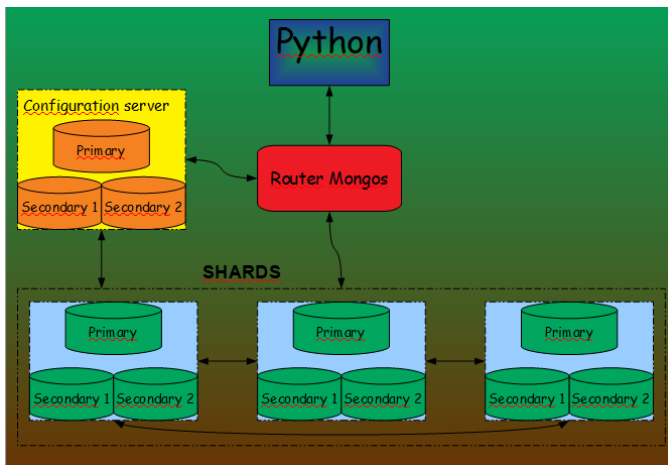


Fig. 1. Struttura del DB partizionato

IV. FLUSSO E STRUTTURA DEI DATI

La fonte principale di dati sono le API di Kiva, che però pongono precise limitazioni al flusso di dati in uscita. In particolare:

- Il formato dei dati esposti da Kiva può essere JSON, HTML ed in alcuni casi XML. Abbiamo scelto di richiedere sempre dati in JSON, dato che Python, mediante la struttura del dizionario, è molto potente nel gestire questo formato;
- ogni richiesta può ottenere in risposta al massimo 100 loans, organizzati in pagine; è stato quindi necessario iterare le richieste sul numero di pagina;
- Ci sono due tipi di richiesta possibile, semplice oppure dettagliata, ma la seconda è possibile solo passando gli ID dei loans richiesti all'API; inoltre, tali ID non sono una successione continua, quindi ad ogni passo è stato necessario prima recuperare l'elenco dei 100 ID dei loans da richiedere, poi recuperare i loans veri e propri utilizzando tali ID;

- Le API hanno dei limiti di richieste al minuto, quindi è stato necessario inserire un tempo di attesa nell'eventualità che una richiesta venga rifiutata per l'eccessivo utilizzo.

Considerati questi limiti, lo script di download dei dati agisce in questo modo:

- Inizializza i parametri cardinali: in particolare, la pagina di loans da cui iniziare a scaricare viene calcolata a partire dal numero di loans già presenti nel database, mentre la pagina di fine viene richiesta alle API;
- Inizia il ciclo di richiesta e salvataggio dei dati. Ogni passo effettua una richiesta per 100 loans, che vengono però scaricati in due passaggi: al primo passaggio richiede gli ID dei loans da scaricare contenuti nella pagina in oggetto, e se tale richiesta va a buon fine, al secondo passaggio invia la richiesta per ottenere i loans veri e propri con gli ID scaricati;
- Effettua la manipolazione dei dati, descritta nella sezione successiva, e salva il blocco di loans sul database.

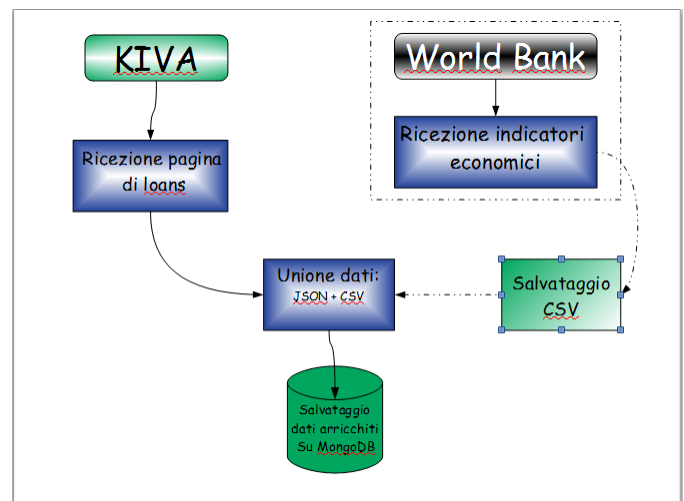


Fig. 2. Flusso di dati

V. MANIPOLAZIONE DEI DATI

Ci sono due livelli di manipolazione dei dati, che verranno chiamati *arricchimento* e *controllo qualità*. Il primo avviene durante il download dei dati, mentre il secondo avviene sul database già formato.

- **Arricchimento:**
Avviene durante il download dei dati, prima quindi dell'inserimento degli stessi nel database. Ad ogni passaggio, lo script scarica 100 documenti JSON contenenti i dati dei loans. Si è scelto di mantenere la struttura dei dati, aggiungendo però alcune informazioni. Al campo "location", contenente appunto informazioni sul paese dove il loan è stato richiesto, sono stati aggiunti i valori dei 4 indicatori economici relativi al paese e all'anno del loan in oggetto.
- **Controllo qualità:**
Avviene sul database già completato per risolvere due gravi problemi dei dati scaricati dalle API: presenza di documenti duplicati, dovuti sia al metodo di salvataggio per pagine che ad errori interni al database di Kiva, ma soprattutto i nomi di alcuni paesi sono diversi dai corrispettivi usati dalla World Bank, rendendo impossibile l'aggiunta degli indicatori in molti casi. Entrambi i problemi sono stati gestiti con degli script Python nel seguente modo:
 - Duplicati: vengono trovati tutti i documenti che comparivano più volte con lo stesso valore nella voce "id". Tali documenti sono poi stati rimossi in blocco e reinseriti ciascuno una volta, così da non perdere dati.
 - Nomi di paese non combacianti: innanzitutto abbiamo confrontato le liste di paesi presenti rispettivamente nel DB di Kiva e nei dati della World Bank, individuando quindi i paesi con nomi differenti. Mediante un'opportuna libreria di Python abbiamo confrontato la distanza tra i nominativi, associando i nomi di paesi con maggiore probabilità di match. A questo punto quindi, usando la mappatura tra nomi così creata, abbiamo potuto riempire i buchi sugli indicatori economici nei documenti afferenti ai paesi in oggetto.

VI. QUERIES EFFETTUATE E RISULTATI

Sono state effettuate 3 queries:

- 1) Quanti loans ci sono e a quanto ammontano in totale per ogni paese?
- 2) Quanti loans vengono chiesti ogni giorno ed ogni ora per un paese esempio (Kenya)?

- 3) Per 8 paesi scelti, c'è correlazione tra gli indici economici e il numero (e/o l'ammontare) dei loans finanziati?

A. Query 1

La prima query restituisce, per ogni paese contenuto: il numero di loans totali, ammontare totale dei loans finanziati in quel paese, ammontare medio dei loans finanziati in quel paese e date del primo e dell'ultimo loan registrato.

Da questa query, alcune informazioni interessanti sono:

- Il paese che in assoluto ha ottenuto la maggior quantità di denaro in loans sono le Filippine, con più di 100 milioni di \$ (per l'esattezza, 100.367.675):
- I paesi col prestito medio più alto sono Bangladesh, Canada e Papua Nuova Guinea, dove è stato emesso un singolo prestito di 50000 \$. Se si eliminano queste situazioni inusuali si ottiene che il paese col prestito medio più alto è Porto Rico, con una media di 6032.64 \$ a prestito.
- I paesi dove Kiva è stata attiva più a lungo sono Cambogia, Honduras, Kenya e Uganda, dove è stata attiva per 4810 giorni (ovvero più di 13 anni, quindi dall'inizio delle attività di Kiva fino ad oggi).

B. Query 2

La seconda query analizza solo il Kenya, scelto come paese esempio. Alcune informazioni che ne abbiamo ricavato sono:

- Il giorno nel quale sono stati erogati il maggior numero di prestiti è il 2 giugno 2011, con 102.912,50 \$ per un totale di 190 prestiti finanziati.
- Il mese nel quale sono stati erogati il maggior numero di prestiti è luglio 2011, con 746.012,08 \$ per 1,253 prestiti finanziati.
- Abbastanza prevedibilmente, l'anno nel quale sono stati erogati il maggior numero di prestiti è il 2011, con 7.943.622,05 \$ per 12.201 prestiti.

La figura 3, visibile nella pagina seguente, mostra l'andamento dell'ammontare dei prestiti erogati in Kenya tramite Kiva mese per mese.

Numero di loans in Kenya dal 01/07/2007 al 31/05/2018

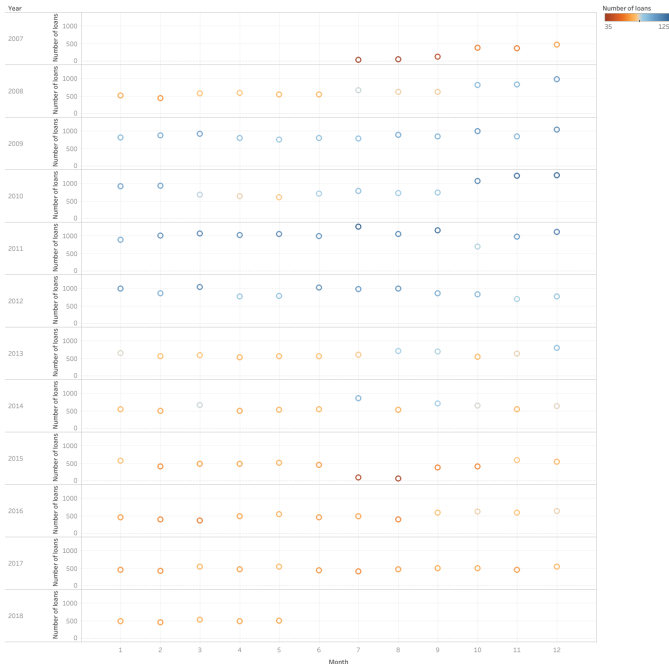


Fig. 3. Ammontare dei prestiti finanziati in Kenya - andamento

C. Query 3

La terza query ha recuperato, per ogni paese salvato nel database, l'ammontare sia totale che medio dei prestiti finanziati negli anni dal 2005 al 2016, oltre ai valori dei 4 indicatori economici ottenuti dalla Banca mondiale. A titolo dimostrativo, abbiamo scelto 8 paesi sui quali abbiamo calcolato, mediante il linguaggio R, la correlazione tra ciascuno dei 4 indicatori e l'ammontare totale dei prestiti finanziati. Gli 8 paesi scelti sono: Cambogia, Ecuador, Honduras, Kenya, Nicaragua, Senegal, Tanzania ed Uganda. Sono stati scelti per tre motivi:

- Distribuzione geografica: i paesi scelti si trovano almeno su due continenti diversi, garantendo che eventuali effetti osservati non siano locali;
- Povertà: tutti i paesi scelti hanno economie deboli e spesso dipendenti da aiuti internazionali, quindi è plausibile che anche un flusso di denaro relativamente piccolo possa avere risultati macroscopici sull'economia;
- tra i paesi che soddisfano i primi due requisiti, sono quelli con la maggior completezza nei dati provenienti dalla World Bank.

Dato che il periodo temporale è molto breve, sono stati inclusi anche i p-value per rendere l'idea dell'affidabilità delle misure effettuate.

• Confronto tra GDP e Microprestiti

Stato	Coefficiente di correlazione	p-value
Cambodia	-0.91292141	0.00034
Ecuador	0.81060692	0.00653
Honduras	0.73435147	0.00653
Kenya	0.89126448	0.00094
Nicaragua	-0.31134692	0.32457
Senegal	-0.71014433	0.00966
Tanzania	-0.00496606	0.98777
Uganda	-0.66055367	0.01937

TABLE II
GDP - CORRELAZIONE

Dalla tabella II si può osservare che vi è sempre alta correlazione, tranne nel caso della Tanzania dove però il p-value altissimo fa dubitare della validità della misura. Occorre anche osservare che le misure sono negative in metà dei casi, a significare una relazione di proporzionalità inversa tra il GDP del paese e l'ammontare totale dei prestiti finanziati in esso. Questa situazione inattesa **potrebbe** indicare che la disponibilità di denaro proveniente dall'estero abbia danneggiato l'economia anziché sostenerla, indicando quindi un ambito di ulteriore ricerca.

• Confronto tra GNI e loans

Stato	Coefficiente di correlazione	p-value
Cambodia	0.64165584	0.0247
Ecuador	0.81300041	0.0013
Honduras	0.79353962	0.0021
Kenya	0.89209435	0.0009
Nicaragua	0.65744159	0.0201
Senegal	0.90662717	0.0004
Tanzania	0.76076236	0.0041
Uganda	0.85776845	0.0003

TABLE III
GNI - CORRELAZIONE

Il secondo indice considerato, il GNI, è specificamente creato per includere i flussi di denaro in entrata ed uscita dovuto ai privati, quindi è legittimo attendersi risultati differenti rispetto all'analisi sul GDP. Dalla tabella III si osserva in effetti come siano scomparse le correlazioni negative ed i p-value siano tutti molto bassi.

Dato che il GNI, presumibilmente, comprende anche il flusso di denaro dovuto a Kiva, la correlazione tra i due parametri non è inattesa.

Tuttavia, di nuovo, la spiegazione potrebbe semplicemente essere la scarsità di dati a disposizione.

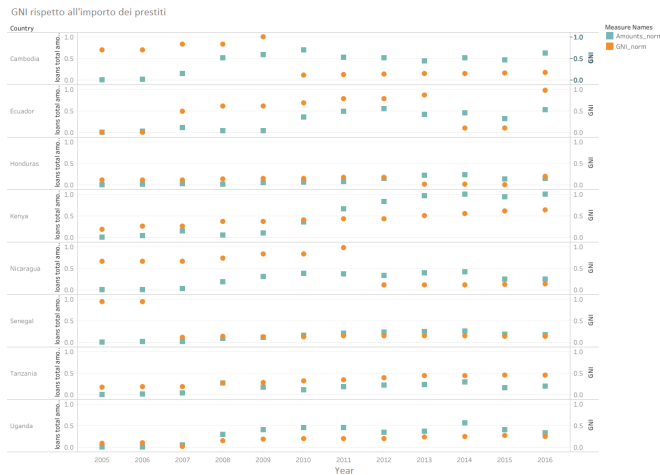


Fig. 4. GNI comparato all'ammontare dei prestiti ricevuti

Può essere più interessante notare dalla figura 4 quanto gli andamenti siano simili, ad indicare che, forse, i loans di Kiva hanno avuto un pesante impatto su economie così deboli.

- Confronto tra GNI-metodo Atlas e loans

Stato	Coefficiente di correlazione	p-value
Cambodia	-0.50172272	0.0965
Ecuador	0.07445345	0.8181
Honduras	0.67178527	0.0167
Kenya	0.89169167	0.0009
Nicaragua	-0.30447916	0.3359
Senegal	-0.59761052	0.0401
Tanzania	0.05240233	0.8715
Uganda	-0.34825085	0.2672

TABLE IV
GNI CONVERTITO CON METODO ATLAS - CORRELAZIONE

Il metodo Atlas considera la media dei tassi di cambio tra la valuta locale e USD sui tre anni precedenti per effettuare la conversione dell'indice. Dalla tabella IV si osserva che le correlazioni negative osservate con il GDP sono tornate, tranne che per la Tanzania (che però mantiene un p-value molto alto). Una possibile spiegazione di questa differenza con il GNI è nella debolezza delle valute di questi paesi, il che conduce a tassi di cambio estremamente variabili e quindi a possibili forti differenze tra il GNI ed il GNI calcolato con metodo Atlas.

- Confronto tra GNI pro capite e Microprestiti

Stato	Coefficiente di correlazione	p-value
Cambodia	0.62232029	0.0306
Ecuador	0.54842373	0.0648
Honduras	0.61675427	0.0326
Kenya	0.11267024	0.7273
Nicaragua	0.55706806	0.0599
Senegal	-0.37631404	0.2279
Tanzania	0.14618904	0.6502
Uganda	0.224072085	0.4838

TABLE V
GNI PER CAPITA - CORRELAZIONE

Il GNI pro capite è il risultato del GNI rapportato alla popolazione in età lavorativa, quindi è una grossolana misura della distribuzione della ricchezza tra la popolazione attiva. Dalla tabella V si osserva come i GNI per capita di Cambogia, Ecuador, Honduras e Nicaragua abbiano correlazione positiva con l'ammontare dei microprestiti e p-value bassi, quindi è ragionevole supporre che i prestiti di Kiva abbiano avuto un effetto benefico in questi stati.

REFERENCES

- [1] Kiva.org [ONLINE]
Available at: <https://www.kiva.org/>
- [2] World Bank [ONLINE]
Available at: <http://www.worldbank.org/>
- [3] Investopedia: What is microcredit [ONLINE]
Available at: <https://www.investopedia.com/articles/personal-finance/040715/what-microlending-and-how-does-it-work.asp>
- [4] Microworld.org: brief history of microcredit [ONLINE]
Available at: <https://www.microworld.org/en/about-microworld/about-microcredit>

APPENDIX

Numero di loans in Kenya dal 01/07/2007 al 31/05/2018



GNI rispetto all'importo dei prestiti



4