



**APES WEEK 2019**

ASIA-PACIFIC ECONOMIC STATISTICS WEEK

Integrating economic statistics in monitoring the 2030 Agenda

**Calculating decile inflation in Iran by using  
online prices data and fuzzy clustering  
method**

**Calculating PPI for the arts, entertainment  
and recreation group of activities by using  
open data**

**Statistical Center of Iran (SCI)**

# Overview

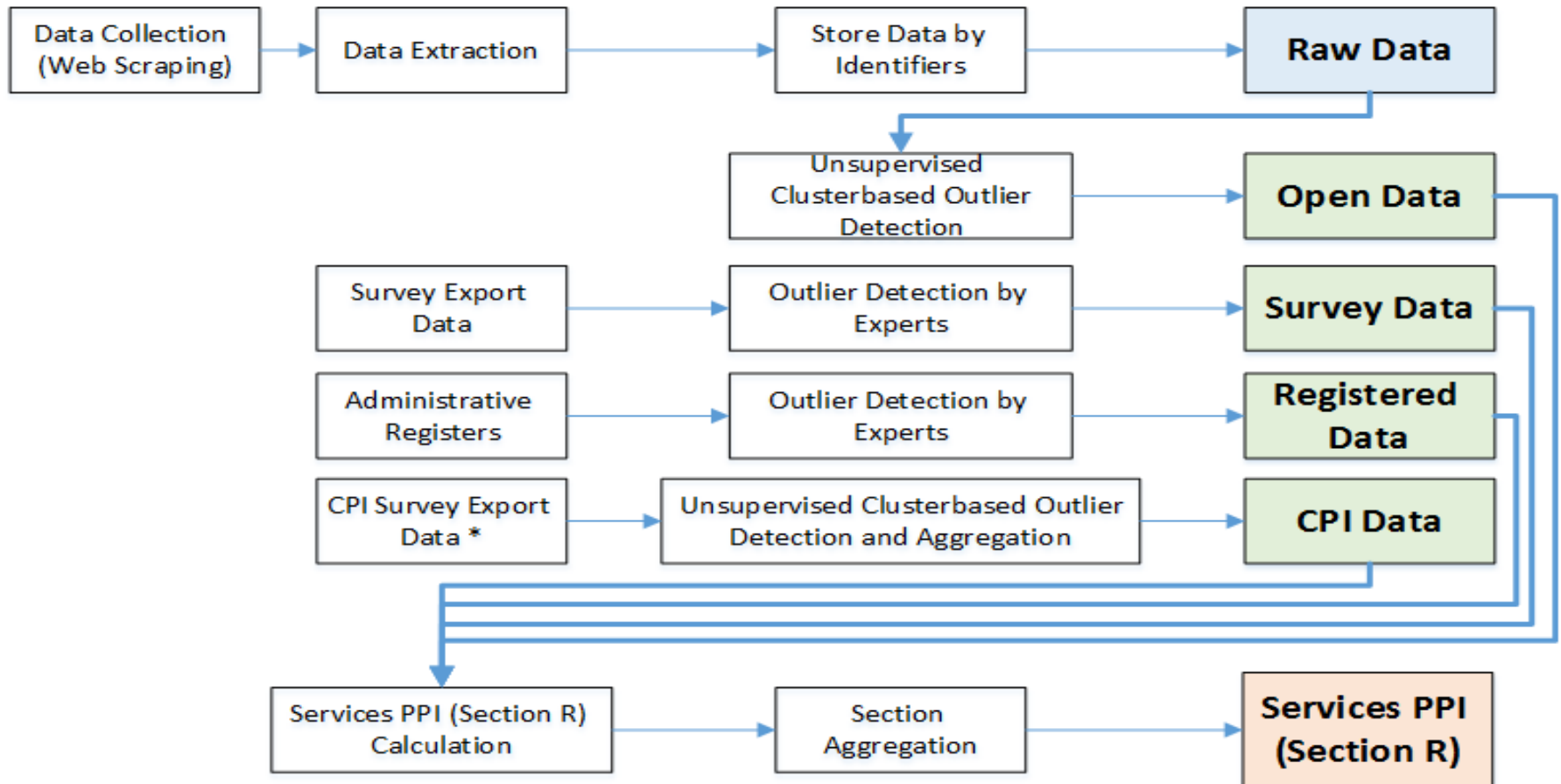
1. Webscraping online data enables price collection which is
  - high frequency
  - easy data collection
  - low cost
  - high accuracy
2. The Fuzzy Cluster Method was used to calculate decile inflation
  - dispersion of price data for consumer goods and services
  - we use fuzzy clustering method to identify the right prices for different income deciles
  - cluster method used to recognise the price paid by each decile
  - fuzzy clustering method used for vague deciles
3. Also tested the use of online data for group R (arts, entertainment and recreation) for PPI

# Fuzzy clustering decile inflation

## Key results

- Comparing the results of this method with the traditional method shows
  - There are differences between them
  - The new method is an improvement
  - In this method the inflation between deciles varies considerably

# PPI R group - data collection



\* Same prices were collected monthly in Consumer Price Index Survey

# PPI R Group

## Key results

- Calculating PPI R Group in the ISIC 4 lowered costs because of
  - Not employing new survey agents
  - Not training survey agents
- Calculating PPI in the ISIC 4 with more sectors

# Fuzzy clustering decile inflation Challenges for production

1. Calculating relative price - changes to index number formulae required (Dutot)
2. In this fuzzy cluster method we don't always have all deciles (when we have low dispersion)
3. There is missing data (when we were faced with changed specification of items)
4. We need ongoing monitoring of which products and services prices are available online

# Using online data in CPI and PPI

## Key challenges

1. New and higher-tech software for collecting, cleansing and classifying information needs trained statisticians and IT engineers.
2. Ongoing software maintenance and redevelopment required due to periodic changes in source sites.
3. Investment for changed production processes will be needed for web scraping
4. Increasing the traffic burden on the source site when collecting price data.
5. Copyright or agreement to use must be coordinated with the custodians of sites.



**APES WEEK 2019**

ASIA-PACIFIC ECONOMIC STATISTICS WEEK

Integrating economic statistics in monitoring the 2030 Agenda



Thank you



# Discussion points

- Changes to production processes require investment
  - software infrastructure
  - different staff expertise
  - monitoring changing raw data
- Product classification is challenging because it requires manual processing
- Methodology (eg index formulae) needs to be reconsidered with this different data
- Volumes sold are needed at some point in the process - either from data providers or household expenditure surveys