

USE OF SCANNER DATA IN CONSUMER PRICE INDEX STATISTICS



Digital Transformation and Projects Department

SESRIC Headquarters - Ankara, 28 September 2020

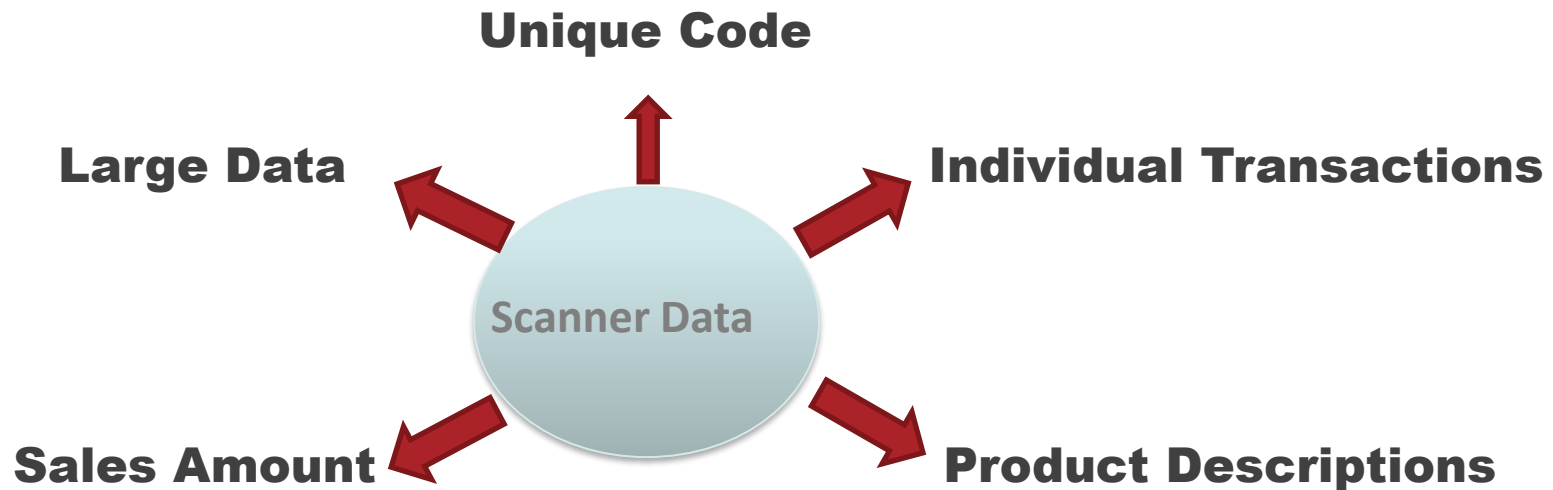
INTRODUCTION

Obtaining Data

Advantages and Challenges

Future Works

Scanner data has become popular in national statistical systems to improve the accuracy of consumer prices index (CPI).



OBTAINING DATA

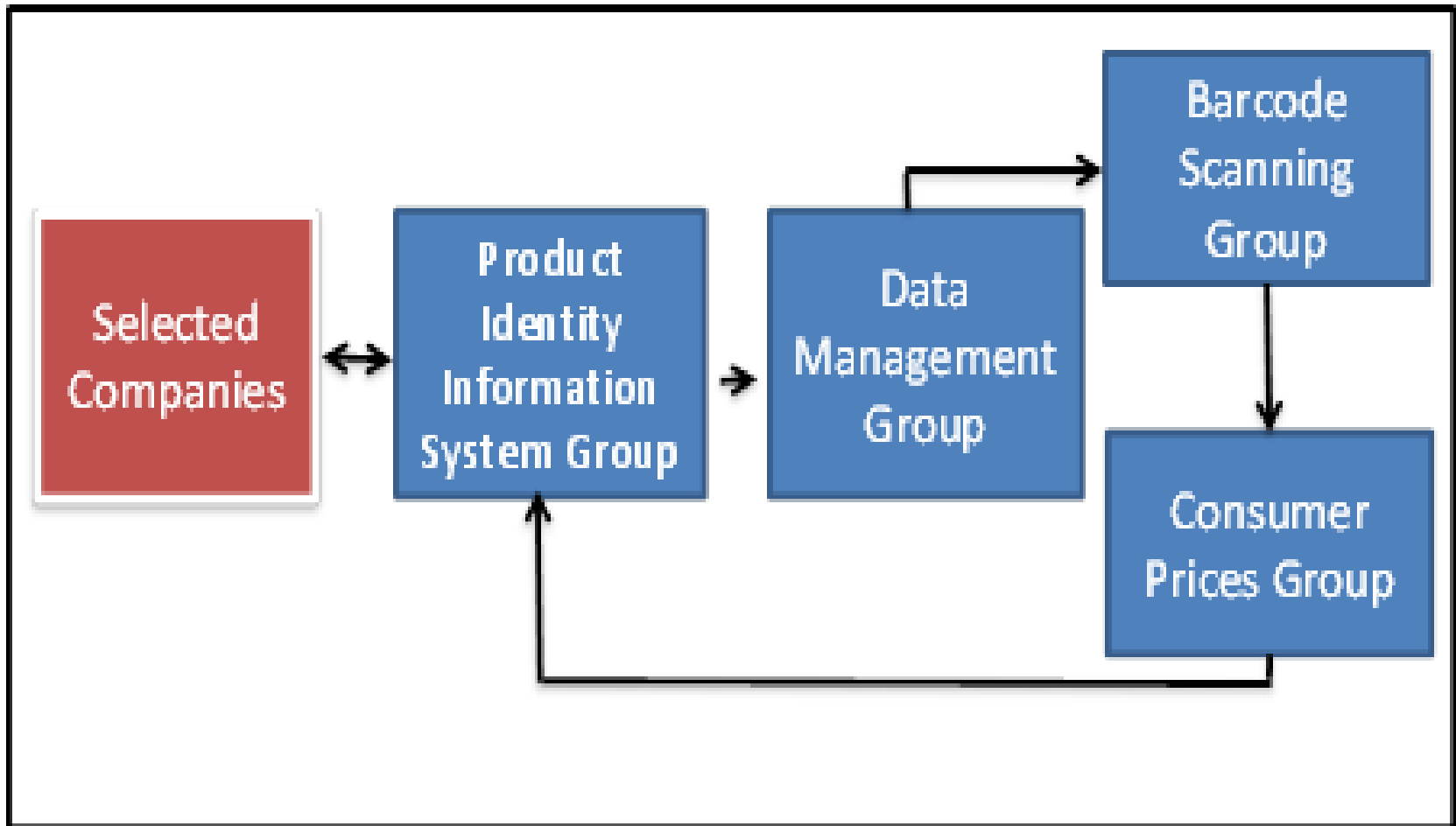
➤ *Who worked for this project?*

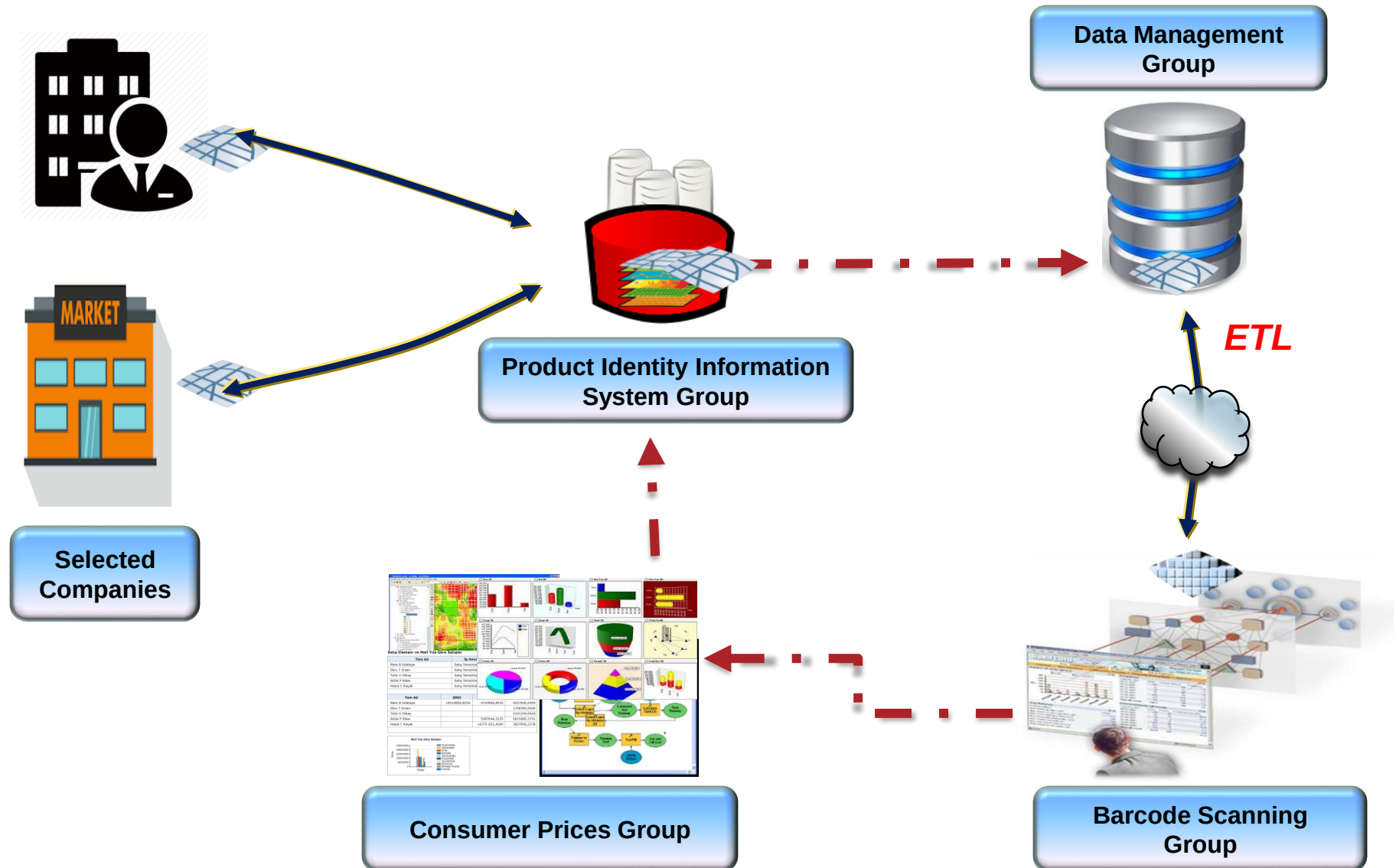
Obtaining comprehensive data sets from different retailers is require expertise and skills in many different fields like communication, statistics, information technologies and law. It is essential to plan and coordinate all the process in which more than one units are involved in the NSO

➤ *How the data was compiled?*

Scanner data can be compiled directly from companies or from intermediary institutions that keep invoice information electronically. In the Turkish Statistical Institute (TURKSTAT) version, data is compiled directly from the companies.

Scanner Data Project in TURKSTAT

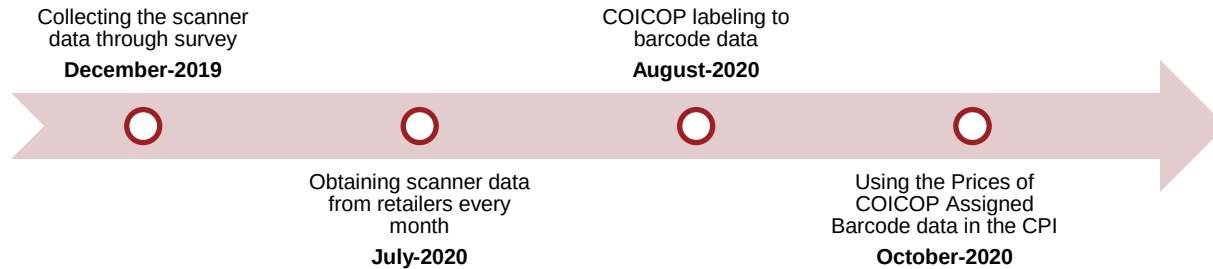




Project Planning

Scope Scanner dataset is compiled from 7 different retailer companies and 1152 divisions

Time



Project Team

•**Data Obtaining Phase**

•Product Identity Information System Group, Data Management Group, Barcode Scanning Data Group

•**Barcode Processing Phase**

•Barcode Scanning Data Group, Consumer Price Group

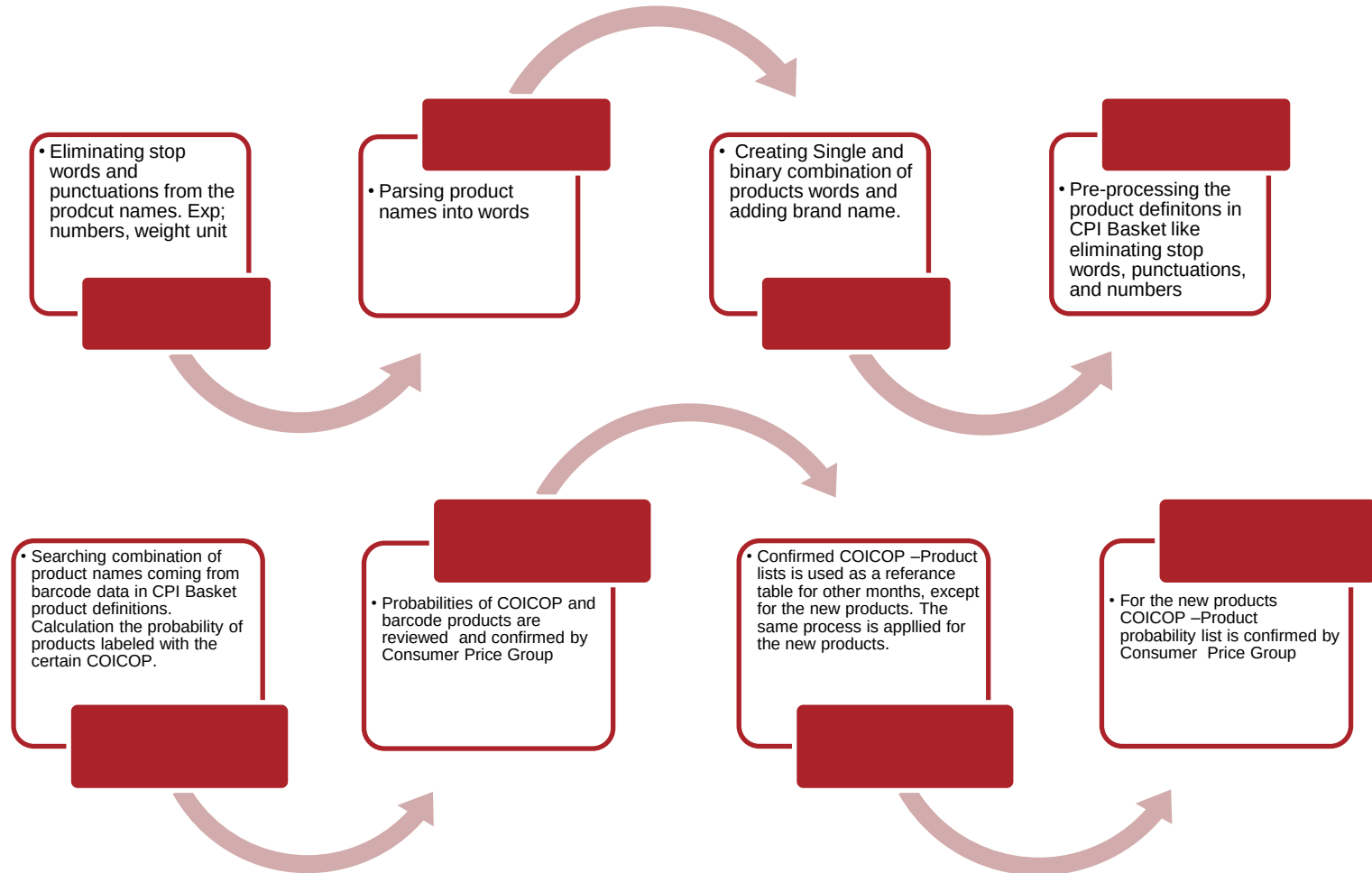
• **CPI Calculation Phase**

•Consumer Price Group, Barcode Scanning Data Group

•Barcode Scanning Data Group : 1 Analyst 1 ETL expert, 2 database developer

Human Resources

COICOP Coding



Price Calculation

- Unit price should be obtained for CPI.
- Barcode prices should be arranged according to the unit of quantity in the definition of COICOP in order to compare the obtained unit prices with the prices of the rifle. (Ex: Napkin 1 piece or 100 pieces, Garbage bag such as 3 meters or 15 pieces)
- Prices must be found by dividing the prices by quantity values. Each item is analyzed by comparing it with its prices in CPI

ADVANTAGES

➤ *Advantages of using scanner data*

- the ability to account for product and outlet substitutions in some categories of consumer spending,
- the ability to produce indexes with greatly reduced sampling error for some item strata,
- the ability to identify new products and quality enhancements,
- and the ability to ensure a more accurate probability sampling of items with fewer manual steps, a sampling procedure known as “disaggregation”.

CHALLENGES

While scanner data sets provide several positive outcomes, there are also several challenges that need to be overcome by NSOs to compile the CPI.



Challenge	How to handle it ?
Exhaustive data	By determining the selection criteria for both sectors and companies, it has started to work with a certain number of companies. Studies will be disseminated over time according to need.
Reluctance of companies	An effective communication strategy is followed, response burden is reduced to the lowest possible level, and demands of companies are answered quickly. Statistics producers who support and encourage a customer-friendly culture that is mindful of respondent concerns are more likely to be rewarded with improved response rates and careful, accurate responses from respondents.

Challenge	How to handle it ?
Difficulty in compiling requested data (Keeping data in different databases, franchises data, etc.)	Companies are given sufficient time to complete these works and technical support is provided to companies in need.
Ensuring data privacy	Privacy Provisions that located in The Statistics Law of Turkey are followed with precision, access authorizations to received data are restricted.

Challenge	How to handle it ?
Failure to transfer data within specified periods	In order to accelerate the processes within the company and to find a source document for responsible personnel changes, protocols are prepared for each company and submitted to the companies before the ultimate data transfer.
Inadequate product classification of companies	Assignment studies of the classification code (E-COICOP) are continued. Codes are assigned manually, by programming or by machine learning techniques

MAIN WORK STEPS

- i. Examination of legal terms
- ii. Clarifying data characteristics
- iii. Selection of companies
- iv. Contacting the companies
- v. Planning and Performing Data Transfers
- vi. Data Editing
- vii. Data Analyze



FUTURE STUDIES

Expanding the coverage

Telecommunications companies, insurance companies, companies operating in the health sector, airline companies and companies operating in the energy market

Web scraping:

Has started with about 90 websites

Thank you for your attention!

Using Big Data Tools and Machine Learning Techniques to Assign Classification of Individual Consumption by Purpose (COICOP) Categories



Digital Transformation and Projects Department

SESRIC Headquarters - Ankara, 28 September 2020

Outline

- Background
 - Definitions
 - Related Works
- Motivation
- Dataset
- Data Preprocessing
- Proposed Methods & Used Technologies
 - Rule-based COICOP Assignment with Apache Spark
 - COICOP Assignment with Machine Learning & Deep Learning Methods
- Results
- Future Works

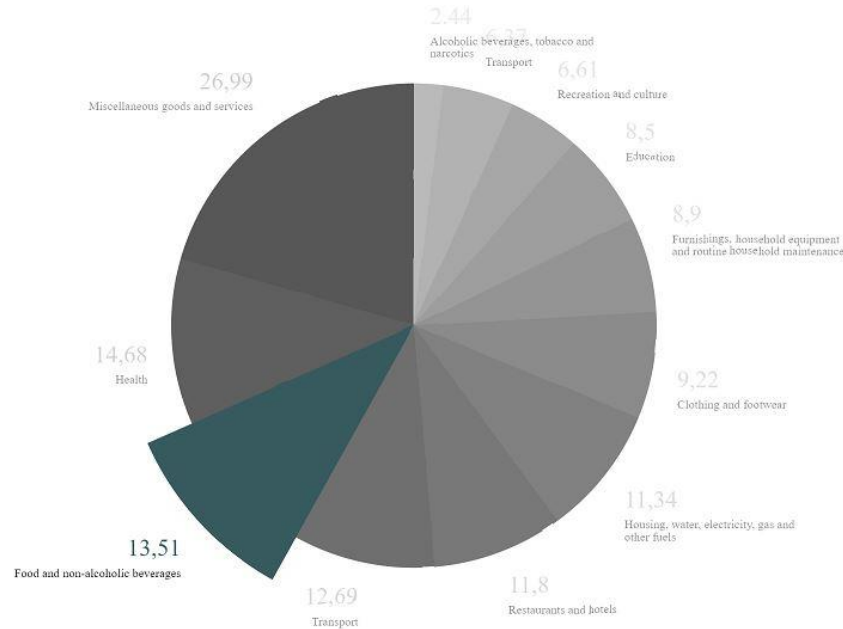
BACKGROUND (Definitions)

➤ *What is CPI?*

CPI (Consumer price index) is a measurement that indicates the average changes in the prices of a certain product and service group purchased by a consumer. Traditionally, CPI is officially calculated with the data of surveys. However, due to the increasing internet usage in recent years, many sources have emerged that can be alternative to survey data.

BACKGROUND (Definitions)

Products must be categorized to calculate the CPI. For this purpose, Classification of Individual Consumption by Purpose(COICOP) is used. 12 main topics have been determined for COICOP. In this study, food category, one of the 12 main topics, was used.



BACKGROUND (Related Works)

In recent years, internet resources have been used widely by statistical offices to calculate CPI. For example, in Argentina, price indices on the PriceStats website have started to be used instead of the official statistics office price indices.



MOTIVATION

The product information obtained is unstructured. In addition, it has a very high data volume due to its high data diversity. Manual classification of products requires high human labor and time. Therefore, it is necessary to prepare a high capacity experiment environment for the analysis and classification of the products.



MOTIVATION

An application with Python programming language has been developed by TURKSTAT to label the data. There are certain rules within this application. Based on these rules, products are labeled COICOP. However, the number of products tagged by this application per day may be limited.

DATA SET

Turkstat collects product information from companies within the scope of the CPI. These products are manually assigned as a 5-digit COICOP code by TURKSTAT branches located in the regions. 152813 individual product names were obtained in the analysis performed on the products assigned a 5 digit COICOP code.

DATA PREPROCESSING

The data set created was used for the machine learning model. However, in order for the data to be used as a training data set, the data must be preprocessed. Numerical values and punctuation marks in product names are removed.

DATA PREPROCESSING

In the final data set, 4000 unique words were obtained. The words obtained have been converted into numerical values so that they can be used in machine learning. In the last stage, 75 percent of the products were made ready for training and 25 percent for testing.

PROPOSED METHODS & USED TECHNOLOGIES

In this study, a number of solutions has been developed by TURKSTAT that estimates the COICOP classification of food products collected from companies. COICOP subcategories are used for classification. This classification consists of a 5 digit code.

PROPOSED METHODS & USED TECHNOLOGIES

We have used Big Data technology for high capacity data processing and analysis. With this method, an increase in performance has been achieved in systems. In addition, the systems have been made more scalable and easier to maintain. Thanks to this technology, an increase in the number of daily COICOP labelable products has been achieved.

We have also used Machine Learning & Deep Learning Technologies to propose an alternative approach to the rule-based COICOP assignation approach.

In this study, Apache Spark, one of the most popular big data technologies, is used. Also, SVM and deep learning techniques such as RNN-LSTM methods are used.

Rule-based COICOP Assignment with Apache Spark

In our rule-based method, Apache Spark, one of the Big Data technologies, was used. The product COICOP assignment application developed by TURKSTAT with Python was re-developed with Apache Spark.

Apache Spark is a Big Data technology tool used for processing large-scale data. In addition, there are components such as machine learning, data grouping and data analysis.

Rule-based COICOP Assignment with Apache Spark

The rules defined in the Python application have been made definable for the Spark application. In our Spark code, user-defined functions (UDF) is implemented so that the rules can be processed for each row in the product dataset. This process was done using Apache Arrow and Pandas UDF tools. The created UDFs were used as parameters in the application, and the product's COICOP code was obtained within the framework of the rules.

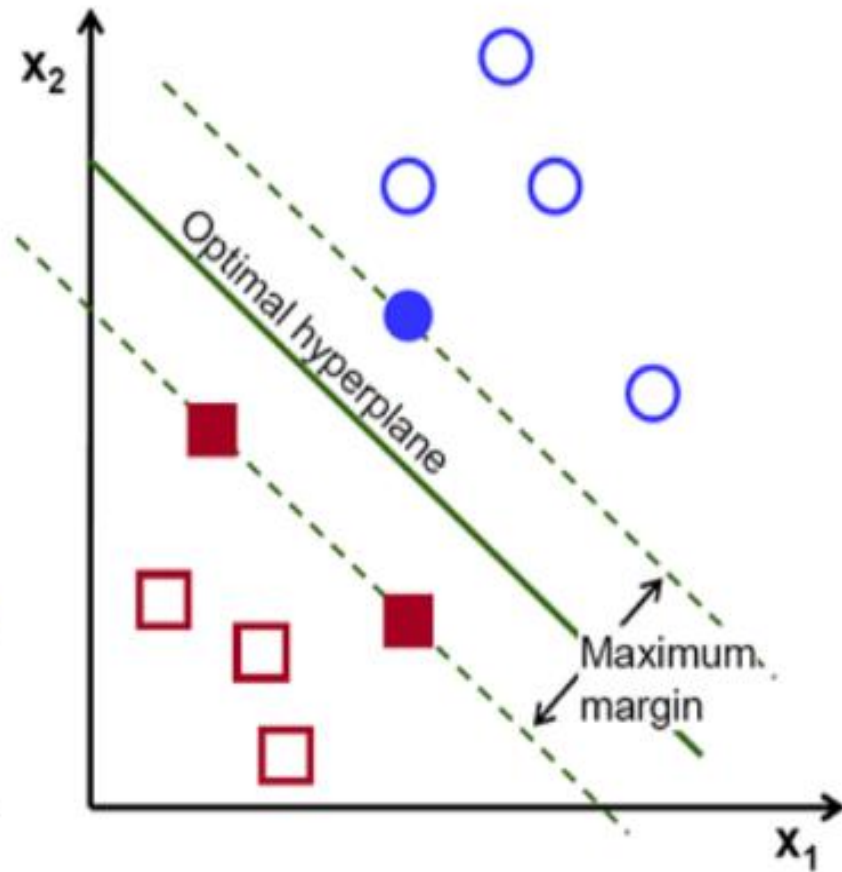
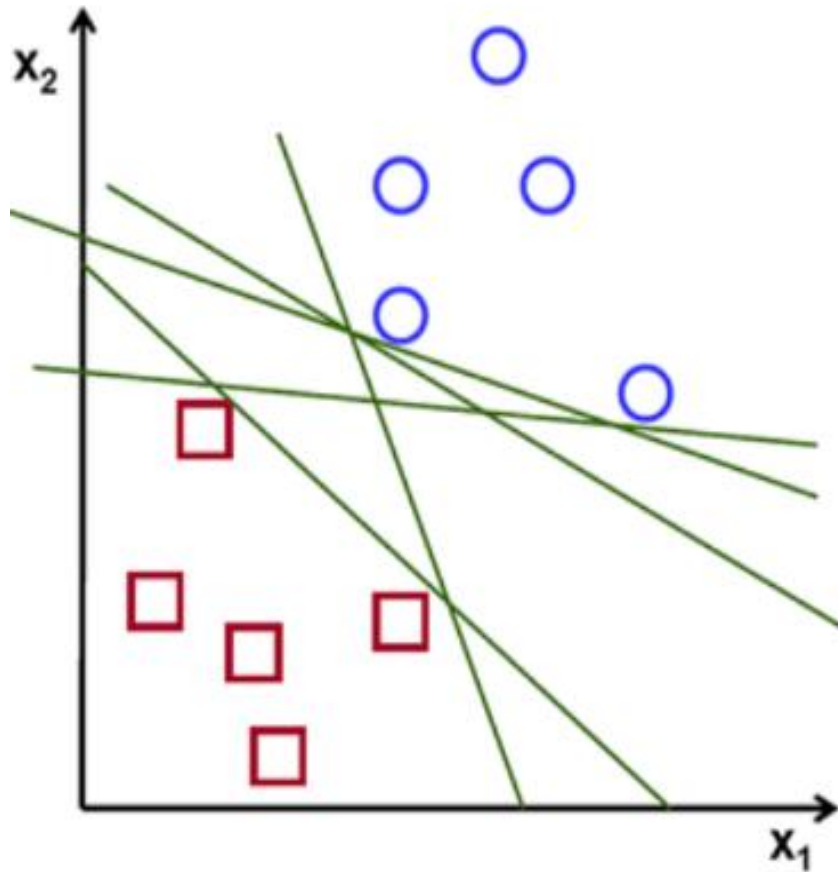
Rule-based COICOP Assignment with Apache Spark

With this new application developed, operations were paralleled and performance was increased 30x. 12 Node Hadoop clustering systems have been installed to parallelize Spark operations.

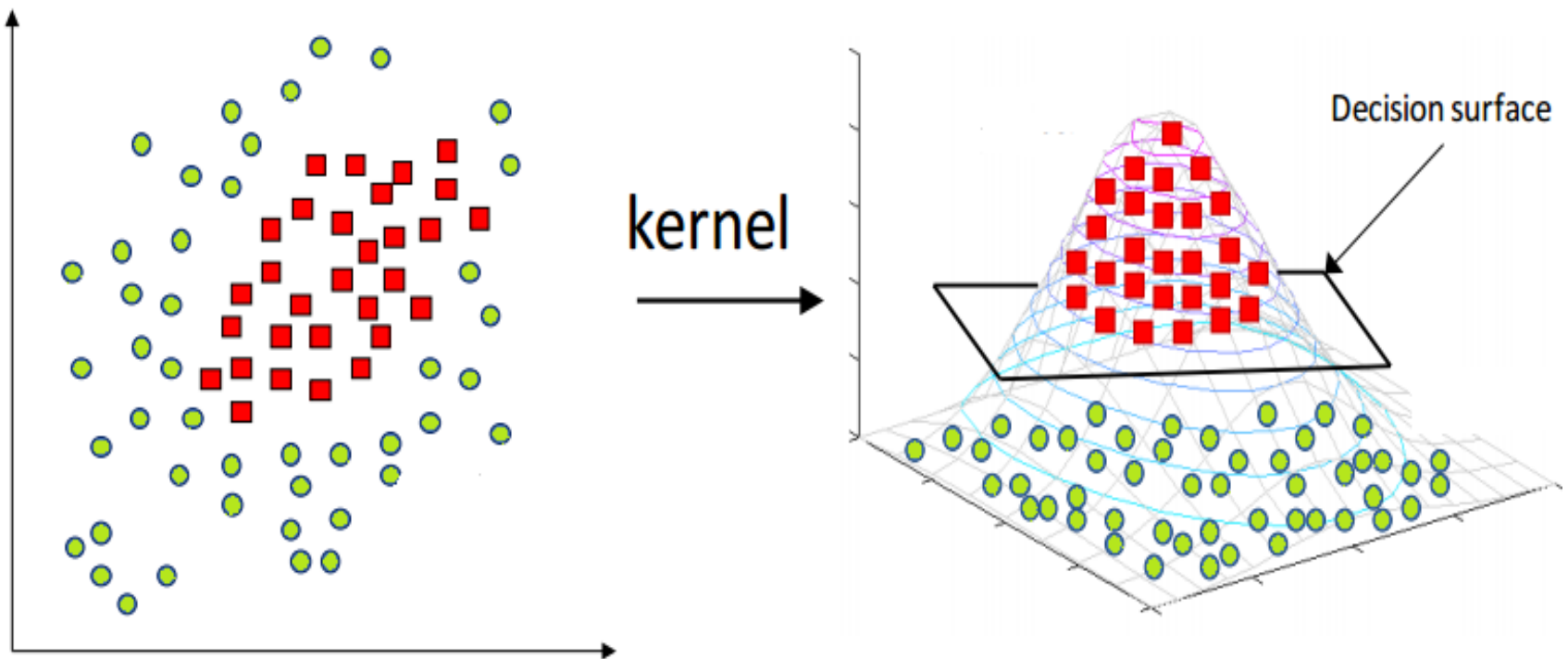
COICOP Assignment with Machine Learning & Deep Learning Methods

By labeling the products, the data set that can be used for machine learning has been made available. SVM (Support Vector Machine), one of the machine learning techniques, was used in this study. 68 COICOP categories were used for this method.

SVM Optimal Decision Function



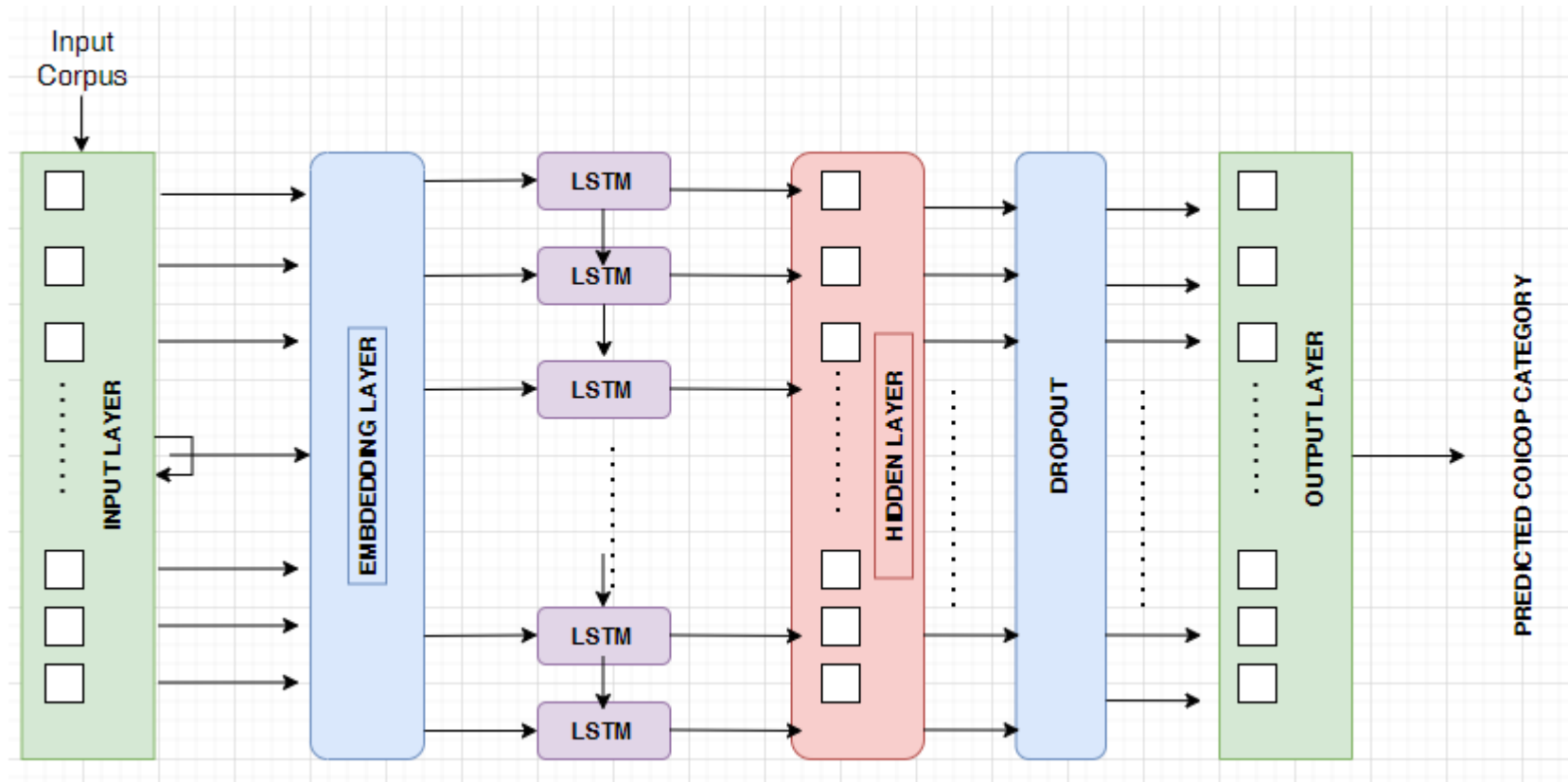
SVM Kernel Trick



COICOP Assignment with Machine Learning & Deep Learning Methods

One of the other machine learning methods used in this study, the RNN-LSTM technique, which is defined as a deep learning model, was used.

LSTM STRUCTURE



RESULTS

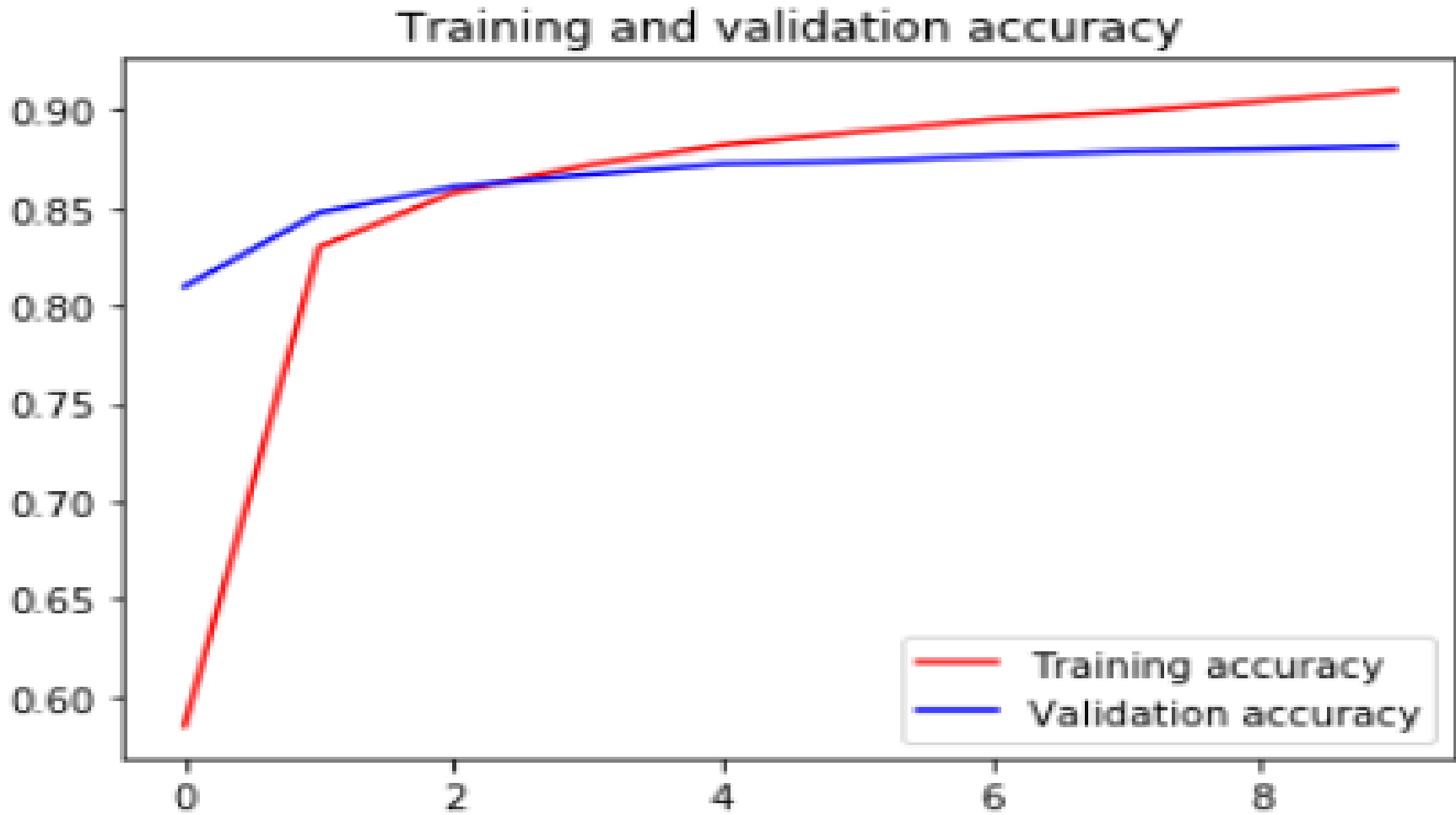
Turkstat has a 16-nodes Hadoop cluster for experimental purposes. Each node in the cluster has 40 core CPU and 380 gb ram. Apache Spark runs on this cluster and it was used for labeling the product data in this work. With this method, the performance of the application has been increased up to 30 times.

In this work, we have also developed machine learning models which were trained with the labeled products. In this approach, no rule was defined and machine learning models learn the ‘rules’ by themselves by examining data. SVM and RNN-LSTM techniques were used for machine learning.

RESULTS

The labeled products were used as a data set for machine learning models. 75 percent of this data set was used for training and 25 percent for testing. In the data set, there are 68 COICOP grain classifications and 152650 product numbers.

The accuracy value of predictions in SVM was 0.868. On the other hand, this value increased to 0.885 in LSTM model.



FUTURE WORK

In future works, images of the products may also be used as an auxiliary data set. This data set will be used for CNN technique, one of the deep learning methods. Thus, with the new model to be created, it is aimed to increase the accuracy rate in estimating the COICOP category of the product.

Thank you for your attendance



TURKSTAT PUSULA PORTFOLIO

Presentation Content

- I. Portfolio Description**
- II. Current & Targeted Situation**
- III. Pilot Project**
- IV. Main Project**

1st Phase

- i. Pusula Analysis Application
- ii. TURKSTAT Calendar
- iii. TURKSTAT Inventory
- iv. Metadata Manamement System

2nd Phase

- i. Effective Use of Data Sources (e.g. Administrative Data)

3rd Phase

- i. Commercialization of TurkStat Pusula Application

TURKSTAT PUSULA PORTFOLIO

PROJECT PHASES

- **1st PHASE:** Automation, standardization and transparency of statistical analysis processes
- **2nd PHASE:** Developing administrative data cleaning tools and diversification operations at administrative data sources
- **3rd PHASE:** Commercialization of TURKSTAT Pusula Project as a packaged program

I. PHASE



II. PHASE

III. PHASE

1st PHASE OF PROJECT: Standardization of Data Process and Analyse Phase

Project Aims:

- Automation, standardization and transparency of statistical analysis processes
- Preparing step by step documents for automated statistical analysis processes
- Allowing statistical production units to focus on more valueable projects
- Determination of «New Project Introduction» (NPI) to be followed for new statistics adaptation to Pusula Application.

Pusula Analysis Application

Through PUSULA Interface application, statistical production units will be able to detect errors, revise, process and analyze data without writing codes

TURKSTAT Calendar

Traceability of everyday flow of organizational processes, meetings, training programs, etc. through a calendar interface.

TURKSTAT Inventory

An inventory of all statistical products TURKSTAT produces and their relation with surveys, administrative data, and also each other.

Metadata Management System

Establishing a metadata management system that consists of Data Dictionaries required for Turkish National Data Pattern, Classification and Code Lists, as well as Variable Bank modules.

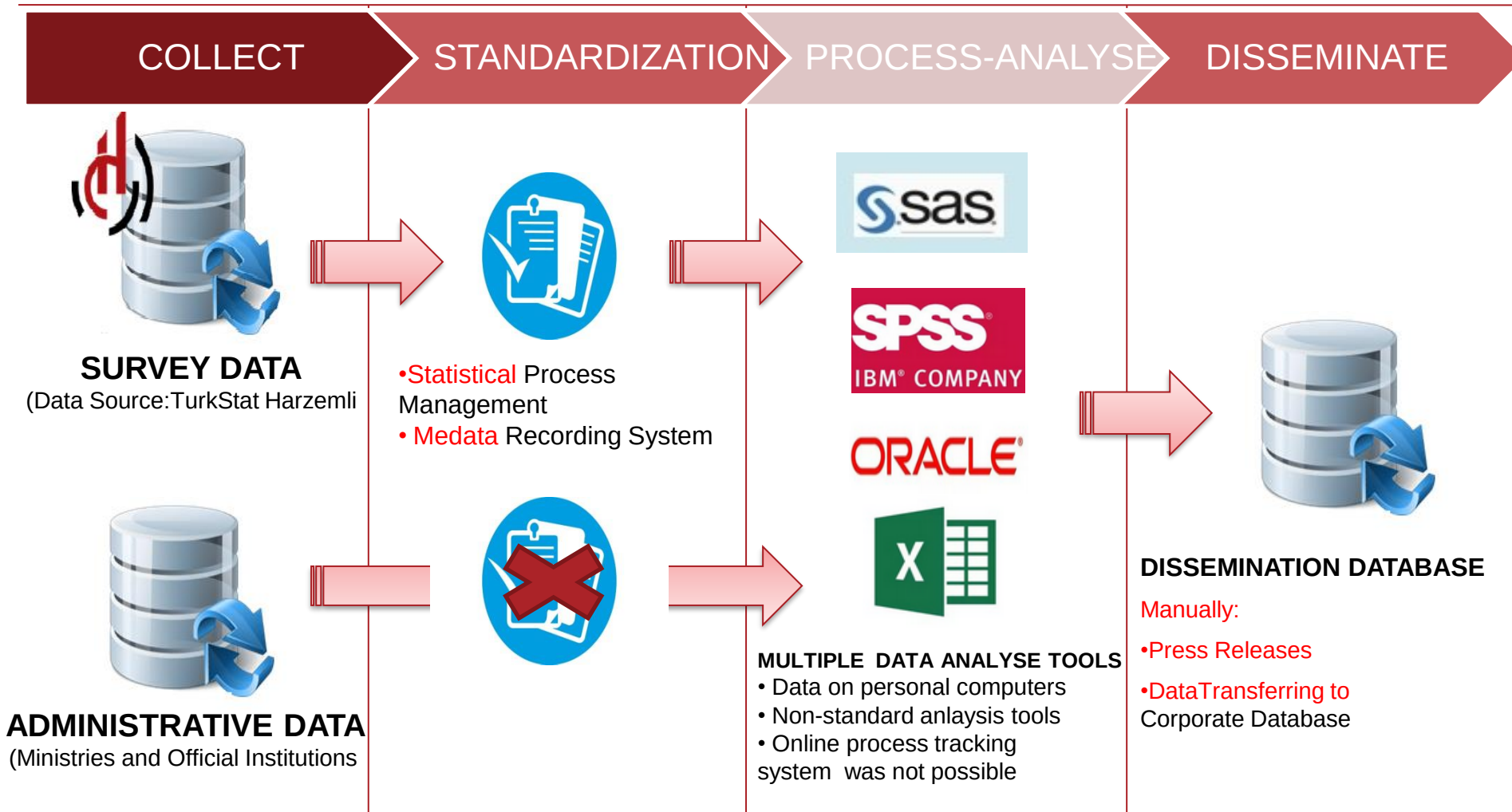
2nd PHASE – Effective Use of Data Sources (Administrative Data)

- National Classification Pattern Project
 - Building a Central Concept Repository
 - Building a Variable Repository
 - Building National and International Classification Code Lists
 - Effective Usage of TurkStat Classification Server
- Preventing Redundant Statistics at National Level and Increasing Statistics Producing From Administrative Data
 - Identification and tracking statistics to be disseminated by Other National Organization and Institution
 - Decreasing respondent burden
 - Decreasing organizational workload
 - Decreasing statistical production burden on other national organization and institutions
- Focusing on Strategic Statistics Production
 - Statistical production of more types of sector and regions
 - Mapping Administrative Data with National Classification Pattern and accelerating new product analysis.

3rd PHASE – Commercialization of TURKSTAT Pusula Application

- Integration of Harzemli Analysis and TURKSTAT Pusula platform
- Making both programs platform independent and packaged programs
- Transition to License Independent (PostgreSQL, Java) open source technologies
- Saving all pl / sql code to a code versioning system (git)

PRODUCTION PROCESS BEFORE PUSULA



TARGETED PRODUCTION PROCESS

COLLECT

STANDARDIZATION

PROCESS-ANALYSE

DISSEMINATE



SURVEY DATA
(Data Source: TurkStat
Harzemli)



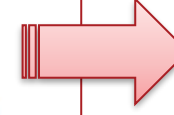
ADMINISTRATIVE DATA
(Ministries and Official Institutions)



- Dynamic Process Management System
- Metadata Management System
- Standard Classification
- National Registration Pattern



**Standard Interface Supported
Data Analysis Software**



DISSEMINATION DATABASE

Otomated:

- Press Releases
- Data Transferring to Corporate Database

Current Situation

Targets

Statistical Production Units produce statistics by **different analysis tools** (SAS, SPSS, Excel, Macro)

All analysis processes are done on **a single tool** (Oracle)

Analysis processes are not documented **in a standard and sustainable manner.**

Sustainable control and documentation of statistical analysis processes integrated into the automation system

Two different production processes (Administrative data and Survey data) are carried out on separate schedules

Development of **Integrated Production Process Scheduling** application

Analysis methods **vary between departments**

Keeping information of all procedures with the **Data Map System**
e.g. Having knowledge on which department uses which method for extreme value analysis

Analysis processes **depends on employees**

Employees can detect and edit data without coding through the interface application.

Current Situation

Unlike survey data, **administrative data** are not generally standardized(**metadata, classification, etc.**).

Insufficient logging on editing data during analyses

Duplicate variables can be produced by repetitive data processing by different units.
(Turnover_Workplace, G_Turnover, Turnover_Indust., Industry_Turnover)

TURKSTAT Statistical Inventory, which consists of 141 statistics, is stored in **Excels**

Targets

Classification of all administrative data according to the standard classification pattern and building infrastructure for the **National Classification Pattern**

With detailed database logging system

Logging of changes made on data, IP, person and time

Monitoring workflows

Identification and revising processes with permanent failures by log analysis

Sharing the data of all variables with other units with the **Data Map System**

Development TURKSAT Statistics Inventory software application

Pilot Project

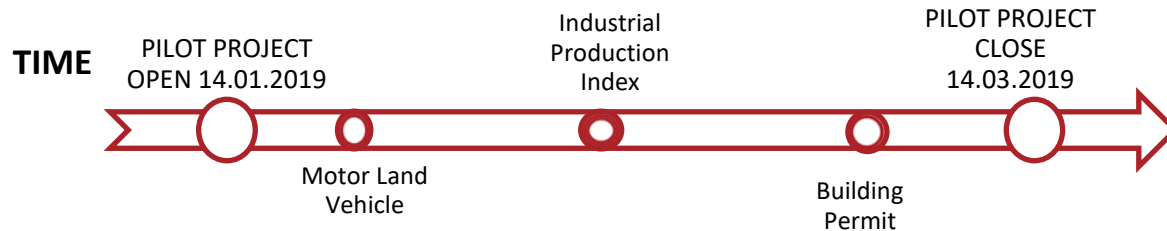
TARGET

- Specifying needs for the main project, management method
- Calculating the total cost, human resources and scheduling by making sample applications

SCOPE

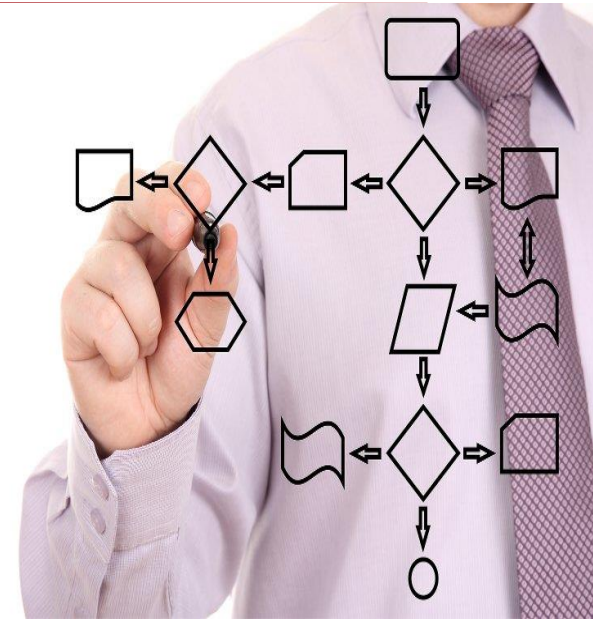
- 3 statistics were selected as pilot projects taking into account data sources.

Statistics Name	Data Source
Motor Land Vehicle Statistics	Administrative Data
Industrial Production Index	Survey +Admin. Data
Building Permit Statistics	Administrative Data

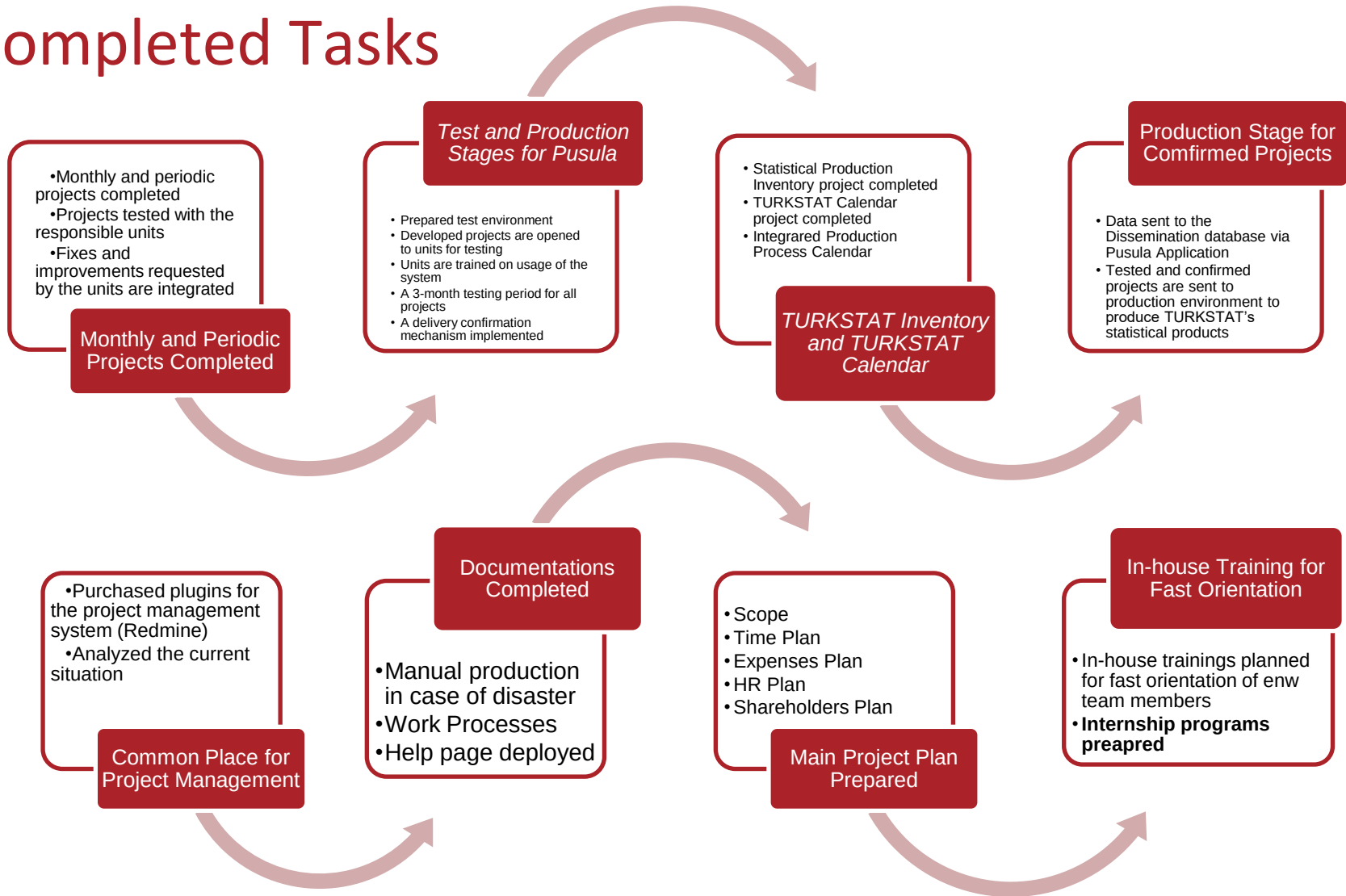


HUMAN RESOURCES

- Project team consisted of 2 administrative and 5 technical experts
- Administrative staff
- 1 Project Manager, 1 Project Manager Assistant
- Technical Staff
- 2 Data Management Specialists, 2 Business Analysts, 1 Software Specialist



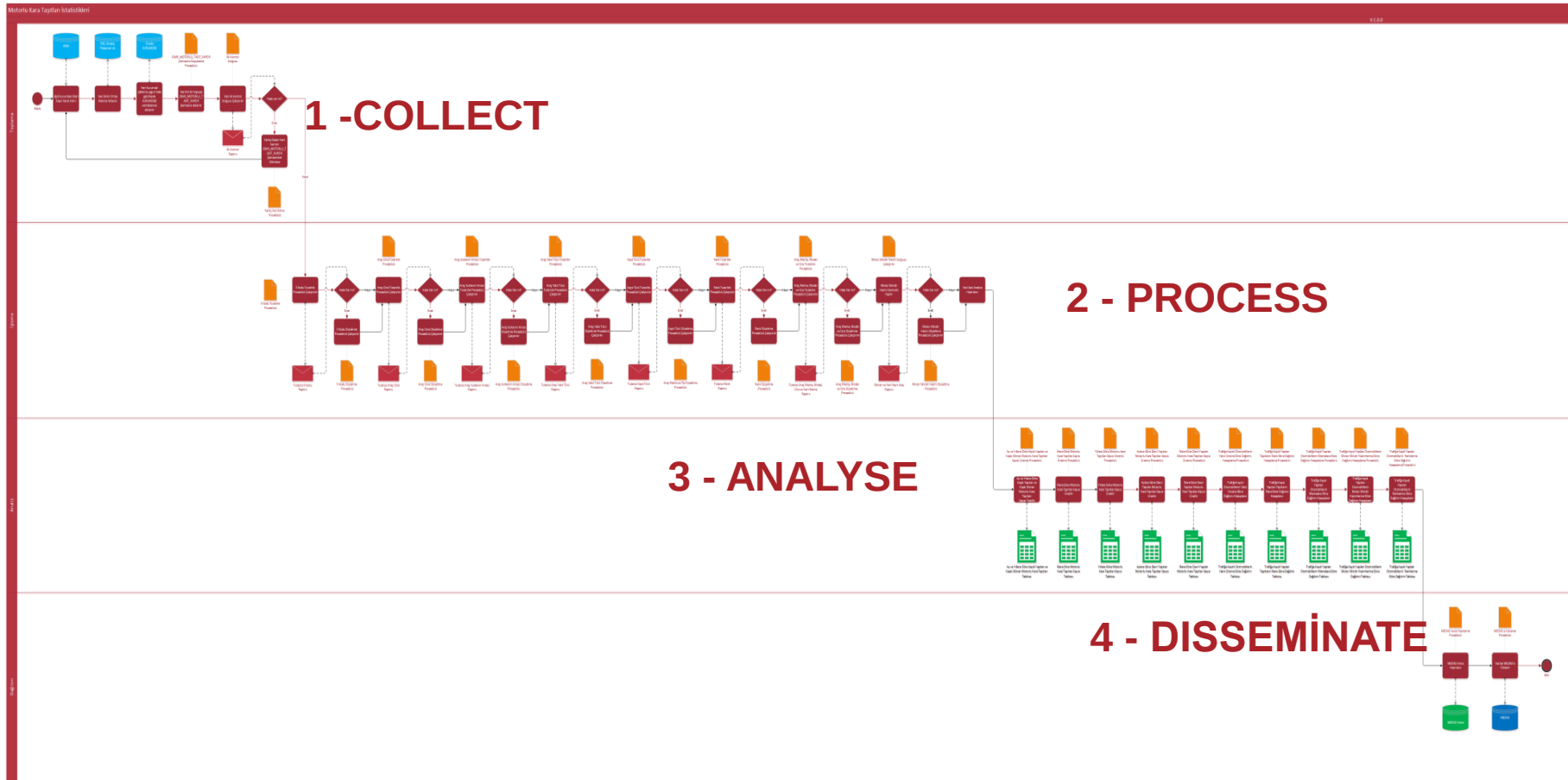
Completed Tasks



Pilot Project Outputs

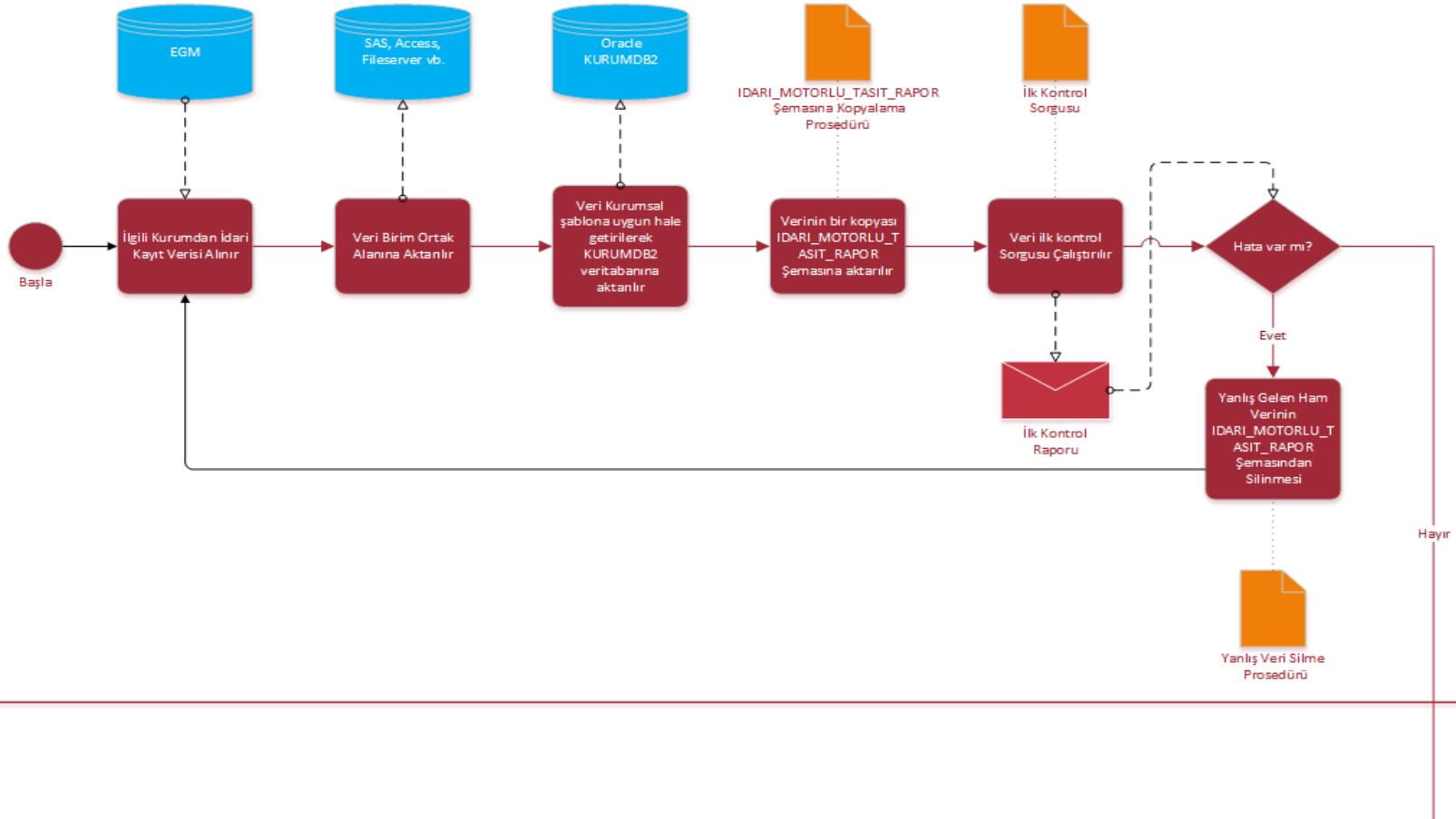
Project Name	Analyse Name	Before	After
Motor Land Vehicle Statistics	<i>Technologies</i>	SAS, Access, Excel, Word, Web Interface	• Oracle, Otomation Interface
	<i>New Outputs</i>	-	• Vehicle Ownership in Households • MTV ile Integration with Motor Vehicles Tax • Automation of subscriber-specific
Industrial Prodcution Statistics	<i>Technologies</i>	SAS, Excel, Word, Web Interface	Oracle, Otomation Interface
	<i>New Outputs</i>	-	• Extreme value detection with Administrative microdata
Building Permit Statistics	<i>Technologies</i>	FoxPro 2.5, Windows95, SAS, Excel	• Oracle, Otomation Interface
	<i>New Outputs</i>	-	• Reducing the burden of information requests by 90% by offering detailed variables to external users

Motor Land Vehicle Statistics- Work Flow

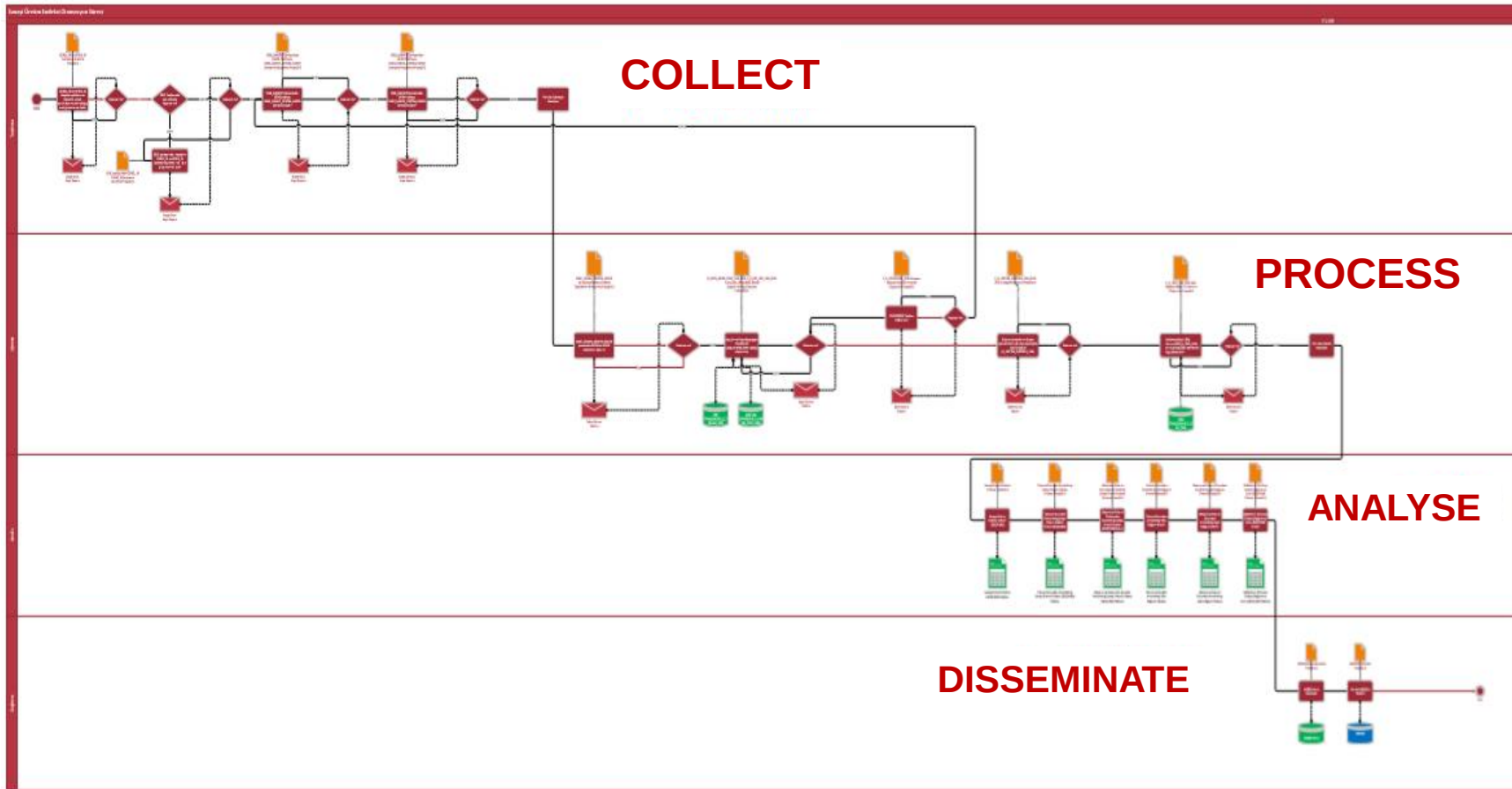


Motor Land Vehicle Statistics- Work Flow

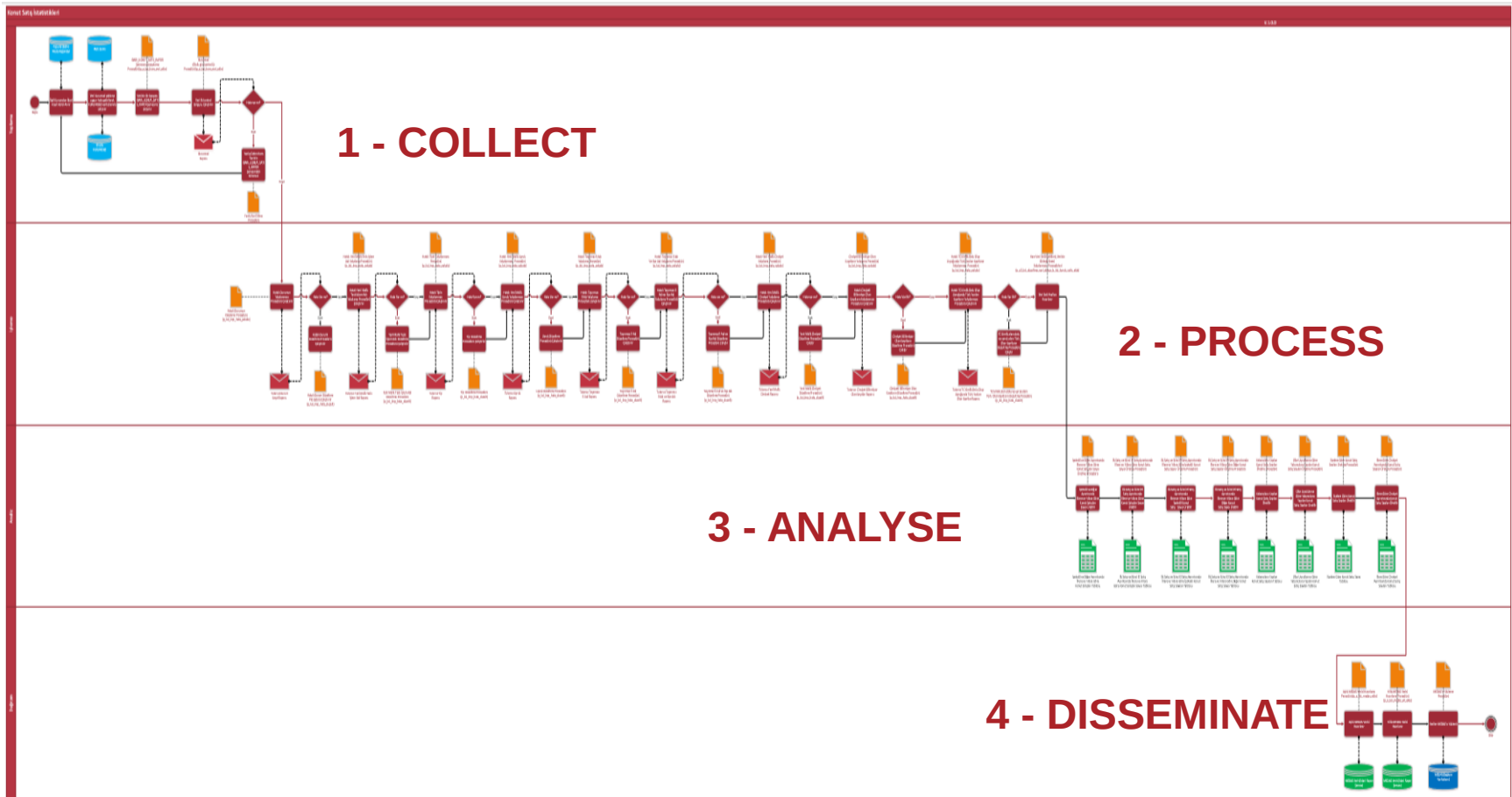
Motorlu Kara Taşıtları İstatistikleri



Industrial Production Statistics- Work-Flow



House Sales Statistics- Work flow



איור PUSULA



126 Çalışma

%44

60

Yapılmadı

8

Sırada

Birim Testinde

1

Yayında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Kullanımında

Haz.20

May.20

[Haber Bültenleri Takvimine Git](#)

24 Haz 2020

24 Haz 2020



24 Haz 2020

25 Haz 2020



29 Haz 2020

29 Haz 2020



PUSULA Automation Software Interface

TUIK PUSULA test | Sektörel Güven Endeksi | Eylül 2019

İşlem Adımları | Çalıştır

İŞLEME | %15

- (YENİ YIL)Örneklem Dosyasının Yükleme
- (YENİ YIL)İK_ANKET_BIRIMI Tablosunun Hazırlanması
- (YENİ YIL)Mektup Adreslerinin Oluşturulması
- (YENİ YIL)İN Dosyalarının Oluşturulması
- İN Atım Tablolarının Oluşturulması**
- İN tablolarının WEBEGİLİM Schemasına Gönderilmesi
- Harzemliden Gelen Anket Verisinin Çekilmesi
- Veri Setinde Kontrollerin Yapılması
- Faktor Tablosunun Oluşturulması
- Faktor Tablosuna Ağırlıkların Aktarılması

Prosedür Çalıştırma Durumları

En Son Çalıştırılan | bugün saat 08:48

- p_sge_insaat_egilim_in_ekle | Tuğba Tuğcu
- p_sge_hizmet_egilim_in_ekle | Tuğba Tuğcu
- p_sge_ticaret_egilim_in_ekle | Tuğba Tuğcu

Önceki Çalıştırılan

bugün saat 08:47

- p_sge_insaat_egilim_in_ekle | Tuğba Tuğcu
- p_sge_hizmet_egilim_in_ekle | Tuğba Tuğcu
- p_sge_ticaret_egilim_in_ekle | Tuğba Tuğcu

05.11.2019

- p_sge_insaat_egilim_in_ekle | Serdal Yolcu
- p_sge_hizmet_egilim_in_ekle | Serdal Yolcu
- p_sge_ticaret_egilim_in_ekle | Serdal Yolcu

04.11.2019

- p_sge_insaat_egilim_in_ekle | Serdal Yolcu

İşlem Başarıyla Tamamlandı | Prosedür çalıştırıldı

İşlem Başarıyla Tamamlandı | Prosedür çalıştırıldı

İşlem Başarıyla Tamamlandı | Prosedür çalıştırıldı

AY	SORU_FORM_KOD	NACE_REV2_4	CEVAPLILIK_DURU
	...	...	...
		41.20	
		43.21	
		41.20	5600
		41.20	3402
		41.20	1600
		41.20	3800
		43.22	3401
		43.33	4100
		41.20	2100
		41.20	3800

Toplam 622 satırdan 1 - 10 arası gösteriliyor | 1 2 3 ... 63

TURKSTAT Calendar

TUIK TAKVİM

[Gündem](#) [Haber Bültenleri](#)

TARİH

22 Haziran 2020

[Dün](#) [Bugün](#) [Yarın](#) [Bu Hafta](#) [Bu Ay](#)

BİRİMLER

[Tüm Birimler](#) [Üretici Birimler](#) [Destek Birimleri](#) [Üst Yönetim](#)

[Başkanlık](#) [Strateji Geliştirme](#)
[Tuiik Başkan Yardımcılığı 1](#) [Tarım İstatistikleri](#)
[Tuiik Başkan Yardımcılığı 2](#) [Ulusal Hesaplar](#)
[Tuiik Başkan Yardımcılığı 3](#) [Uluslararası Ticaret İstatistikleri](#)
[Tuiik Başkan Yardımcılığı 4](#) [Uluslararası İlişkiler](#)
[Hukuk Müşavirliği](#) [Veri Toplama](#)
[Bilişim Teknolojileri](#) [Yazılım](#)
[Demografi İstatistikleri](#) [Yıllık İş İstatistikleri](#)
[Destek Hizmetleri](#) [Yöntem Araştırmaları](#)
[Dijital Dönüşüm Ve Projeler](#) [Çevre Ve Sürdürülebilir Kalkınma İstatistikleri](#)
[Eğitim](#) [İşgücü Ve Yaşam Koşulları](#)
[Fiyat İstatistikleri](#) [Basın Ve Halkla İlişkiler Müşavirliği](#)
[Kısa Dönemli İş İstatistikleri](#) [Özel Büro Müdürlüğü](#)
[Metaveri Ve Kayıt Sistemleri](#) [Veri Kalite Kontrol Kurulu](#)
[Personel](#) [Yayın Dağıtım Kurulu](#)
[Sektörel İstatistikler](#) [Kurullar](#)
[Sosyal İstatistikler](#)

İŞ TİPLERİ

[Üretim Süreci İşleri](#) [İdari ve Mali İşler](#) [Eğitimler](#) [Toplantılar](#) [Yurt Dışı Faaliyetleri](#) [Bilişim İşleri](#) [Diğer](#)

Araştırma Ve Anket Tasarımları Grup Başkanlığı

ÜRETİM SÜRECİ İŞLERİ 1 adet

Ağırlıklı ÖATD Başkanlığı Tarafından Uygulayıcı Daire Başkanlığına İletilmesi

[Hanehalkı bütçe araştırması](#)

2019 Dönem 1

Nüfus Nitelikleri Grup Başkanlığı

ÜRETİM SÜRECİ İŞLERİ 1 adet

Merkez Tarafından Analiz ve Kontrollerin Tamamlanması

[Dünya nüfus günü](#)

2020 Dönem Yıllık

Sağlık Ve Sosyal Koruma İstatistikleri Grup Başkanlığı

ÜRETİM SÜRECİ İŞLERİ 1 adet

Merkez Tarafından Analiz ve Kontrollerin Tamamlanması

[Türkiye sağlık araştırması](#)

2019 Dönem -

Yapısal İş İstatistikleri Grup Başkanlığı

ÜRETİM SÜRECİ İŞLERİ 1 adet

Uygulamanın Altlık Verisinin Oluşturulması ve Veri Tabanına Atılması/Bölge Müdürlüklerine Gönderilmesi

[Yıllık sanayi ve hizmet istatistikleri yatırım harc...](#)

2019 Dönem -

PUSULA Submission Consent

TÜİK PUSULA test MÜGE NAZ YILDIRIM

ÇALIŞMALAR

Katı Yakıt İstatistikleri Aralık Ocak Şubat

Aylık Süt ve Süt Ürünleri İstatistikleri Ocak Şubat Mart

Aralık Ocak Şubat

Aralık Ocak Şubat

Ocak Şubat Mart

Ocak Şubat Mart

Mart Nisan Mayıs

Şubat Mart Nisan

ÖNEMLİ

KISA DÖNEMLİ İŞ İSTATİSTİKLERİ DAİRE BAŞKANLIĞI'na bağlı CİRO VE HACİM GÖSTERGELERİ GRUBU tarafından yürütülmekte olan **Perakende Satış Endeksleri** çalışmasının ikinci kontrolleri TÜİK PUSULA arayüzü üzerinden test edilmiştir. Birimimiz tarafından farklı analiz araçları ile yapılan hesaplama sonuçları, yine birimimiz tarafından TÜİK PUSULA üzerinden yapılan hesaplama sonuçları arasında herhangi bir farklılığa rastlanılmamıştır.

☒ Onaylanan test aşaması üzerinde geri alma, yineleme ve yeniden uygulama işlemleri yapılamaz.

Onayla Vazgeç

Sürüm: 0.22.1-test

PUSULA Help Page

[Anasayfa](#)[PUSULA Nedir?](#)[PUSULA Kullanım Kılavuzu](#)[PUSULA Yardım](#)[Sık Sorulan Sorular](#)

Pusula Yardım Platformu

[ARA](#)

PUSULA YARDIM PLATFORMU

Ekonomik Güven Endeksleri

EGE- Süreç Dokümanı

EGE-Kullanıcı Yardım Dokümanı

[View all 2 articles](#)

Sanayi Üretim İstatistikleri

SUE- Kullanıcı Yardım Dokümanı

SUE- Süreç Dokümanı

[View all 2 articles](#)

Motorlu Kara Taşıtları İstatistikleri

Motorlu Kara Taşıtları İstatistikleri

[View 1 article](#)

Tüketici Güven Endeksleri

TGE- Kullanıcı Yardım Dokümanı

TGE- Süreç Dokümanı

[View all 2 articles](#)

Sektörel Güven Endeksi

SGE- Kullanıcı Yardım Dokümanı

SGE-Süreç Dokümanı

[View all 2 articles](#)

Vatandaş Giriş Araştırması

Vatandaş Giriş Araştırması

[View 1 article](#)

Konut Satış İstatistikleri

Konut Satış İstatistikleri

[View 1 article](#)

Finansal Yatırım Araçlarının Reel Getiri Oranları

Finansal Yatırım Araçlarının Reel Getiri Oranları

[View 1 article](#)

MAIN PROJECT

TARGETS

- Automation, standardization and transparency of statistical analysis processes
- Creating an opportunity for statistical production units to focus on value added businesses
- Commercialization of the product as a package program in integration with TÜİK Harzemli Platform
 - Creating the Integrated Data Map
 - Establishing of metrics, such as timeliness of the products, human resources, use of technology, etc.
 - Establishing the Integrated Statistics Process Schedule
 - Creation of official statistics database



KAPSAM

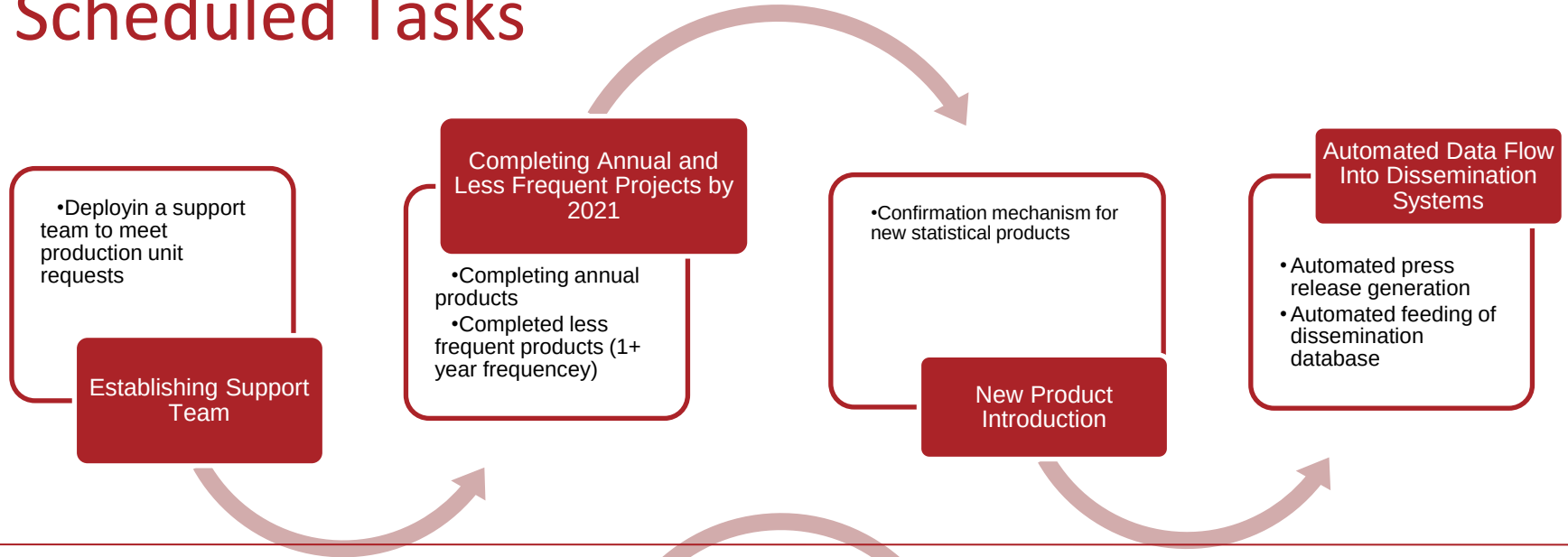
- 141 statistics are actively being produced in the TURKSTAT Statistics Inventory.

Departments	Monthly	Quarterly	Annually	1 Year +	Others
Price Statistics	6		8		4
Short-term Business Statistics	6		2	1	1
Tarım İstatistikleri	2	2	7		3
Environment and Sustainable Development Indicators			3	10	
Labor Force and Living Conditions	3		5		
Social Statistics			14	1	
Annual Business Statistics			4	1	2
National Accounts		1	2	2	
International Trade Statistics	2	3	3		
Demography Statistics			13		1
Sectoral Statistics	3	2	8		
Total	22	8	69	15	10

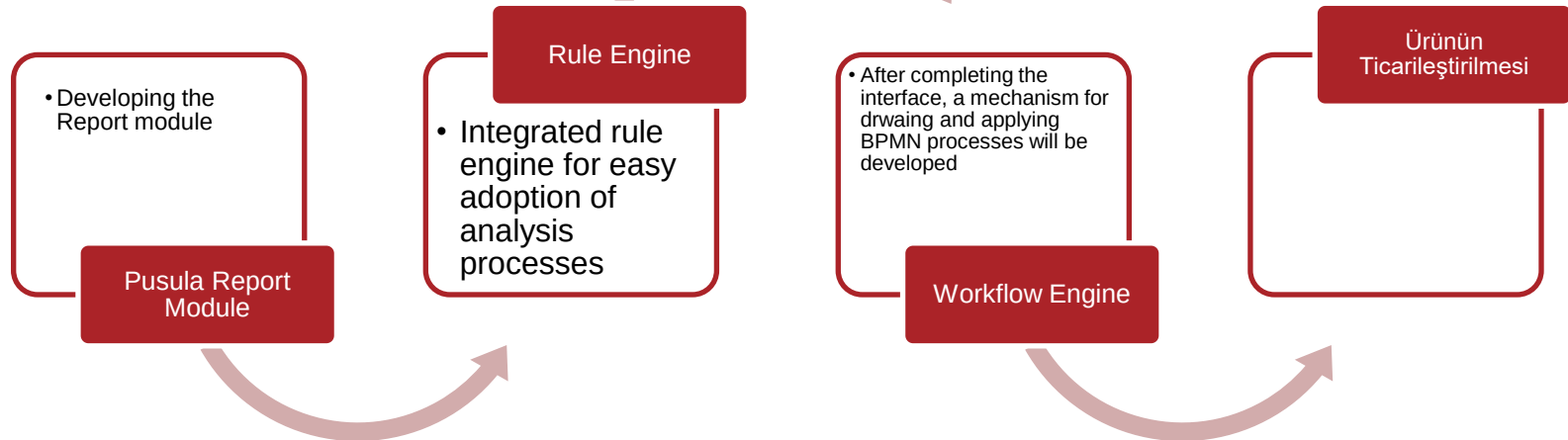


Scheduled Tasks

VERİ YÖNETİMİ



YAZILIM



Planning Schedule– I. PHASE

DATA MANAGEMENT

PROJET
START

18.03.2019

Annual
Statistics

• 75 Statistics

15.10.2020

PROJECT
END

31.12.2020

31.12.2019

Monthly – Quarterly Statistics

- 30 statistics-Monthly
- 8 statistics-Quarterly

15.12.2020

+1 Year

- 14 statistics

SOFTWARE

PROJET
START

08.03.2019

Building
Integrated
Data Map
(Survey + Admin.
Data)

01.10.2019

Work Flow
EngineSoft
ware

30.11.2019

Commercialization
of the Product

15.12.2020

30.06.2019
Completing
Interface
Software

15.10.2019
Rules Engine
Software

30.06.2020
Metadata
Managemnt
System

PROJECT
END

31.12.2020

Human Resource Planning– I. Phase

Dates				Scope	Time	Cost			HR
		Project Start Date	Project Close Date	Overall Project Percents	Workdays	Total man*day	Avg. man*day per Project	TL-Cost	Personnel Needs
	Completed	14.03.19		47%	389	2.957	51	1.612.909	
2020	Remained (Month-Period)		31.12.20	53%	83	3334			40
2021			31.12.2021		344	3334			10

- Cost per statistics adaptation is 51 man*days.
- In order for the project to be completed , which consists of 110 products, until 31.12.2020, 6 technical personnel are required, apart from the existing 10 technical personnel.

Scheduling Preview– I. PHASE

PUSULA Genel Proje Planlayıcısı

Sağ taraftan vurgulananacak bir dönem seçin. Grafik aşağıya gösterge, bu hücrenin yanında verilmiş tir.
Mevcut hafta: 34

ETKİNLİK

Mevcut Hafta : 34

Plan Süresi Filili Başlangıç % Tamamlandı Filili (planın ötesinde) % Tamamlandı (planın ötesinde)

PLAN BAŞLANGICI PLAN SÜRESİ Filili BAŞLANGIÇ Filili SÜRE TAMAMLANMA YÜZDESİ OCAK ŞUBAT MART NİSAN MAYIS HAZİRAN TEMMUZ AĞUSTOS EYLÜL EKİM KASIM ARALIK

ETKİNLİK	PLAN BAŞLANGICI	PLAN SÜRESİ	Filili BAŞLANGIÇ	Filili SÜRE	TAMAMLANMA YÜZDESİ	OCAK	ŞUBAT	MART	NİSAN	MAYIS	HAZİRAN	TEMMUZ	AĞUSTOS	EYLÜL	EKİM	KASIM	ARALIK
PUSULA Pilot Proje Uygulaması (Hazırlık)	07/01/2019	Pilot proje için gerekli olabilecek ihtiyaçların tespiti, yönetim usulünün belirlenmesi ve numune uygulamalar yapılması ile toplam maliyet, insan kaynağı ve zaman planlarının çıkarılması	2	2	2	100%											
	21/01/2019	Pilot projelerin belirlenmesi (Motorlu Kararıştırmalar, Sanayi Üretim Endeksi, Yapı İzin İstatistikleri) ve ön analiz safhalarının tamamlanması (paydaş görüşmesi, ihtiyaç analizi, mevcut dokümantasyonların temini v.b.)	3	6	3	100%											
	21/01/2019	Kullanılacak yazılım mimarisi (Veri Tabanı: Oracle, Yazılım Dili: PL/SQL, Ön Yüz: ReactJS) ve proje yönetim sistemine (KANBAN) karar verilmesi	4	1	5	100%											
	14/03/2019	3 adet pilot projeye ait; süreç akışı, kod listeleri, kullanım kılavuzu ve maliyet dokümanlarının hazırlanması	4	6	5	8	100%										
	14/03/2019	Kodlama ve test süreçlerinin tamamlanması	4	6	5	8	100%										
PUSULA Veri Analiz ve Standardizasyon Projesi	14/03/2019	3 adet Pilot proje kapsamında elde edilen maliyet, insan kaynağı ve zaman verileri üzerinden, TÜİK Ürün Envanterinde yer alan diğer 141 adet ana proje için projeksiyonların yapılması	3	6	3	100%											
	31/08/2019	Aylık üretilen istatistik ürünlerin (30 adet) veri analiz ve standardizasyon kodlamalarının tamamlanması	12	24	11	18	49%										
	31/12/2019	Dönemsel üretilen istatistik ürünlerin (8 adet) veri analiz ve standardizasyon kodlamalarının tamamlanması	36	16	-	-	8%										
	15/10/2020	Yıllık üretilen istatistik ürünlerin (75 adet) veri analiz ve standardizasyon kodlamalarının tamamlanması	52	44	-	-	0%										
	31/12/2020	Bir yıl üzeri vadede üretilen istatistik ürünlerin (28 adet) veri analiz ve standardizasyon kodlamalarının tamamlanması	96	30	-	-	0%										
PUSULA Arayüz Geliştirme	04/02/2019	İstatistik analiz işlemlerinin herhangi bir kodlamaya girilmeden, arayüz üzerinden yürütülmesinin ve loglanmasının sağlanması için ihtiyaç analizi yapılması	6	2	5	2	100%										
	24/06/2019	Tüm projelerde kullanılabilecek tek bir jenerik arayüz tasarım modeline geçilmesi ve kodlamanın başlatılması	8	18	7	18	100%										
	24/06/2019	Test-iyileştirme-test usulü süreli proje testlerinin yapılması	8	18	7	18	90%										
	24/06/2019	Tamamlanan PUSULA projelerinin arayüze aktarılması	8	18	7	18	100%										
	30/06/2019	Arayüzün canlı kullanıma alınması	26	2	-	-	0%										
PUSULA Ürün Envanteri Projesi	24/03/2019	TÜİK tarafından üretilen devam eden istatistik ürünlerin ve bu ürünlerin ait metaverilerin (tanım, üretim sıklığı, coğrafi düzey, veri tipi, kaynak kurum, veri aktarım türü, referans yılı v.b.) tespiti ve pilot veritabanının oluşturulması	13	2	13	2	100%										
	11/04/2019	TÜİK ürünlerine ait envanter ve üretim süreci bilgilerinin izlenebilmesini sağlayacak pilot arayüz/ iş zekası aracının envanter veri tabanı ile entegrasyonunun sağlanması	15	2	15	2	100%										
	10/10/2019	Ürün envanteri projesinin canlı sistemlere aktarılması	32	12	32	-	0%										
Kural Editörü	30/12/2019	Tamamlanan PUSULA projelerinin arayüze aktarılması	36	8	-	-	0%										
	30/12/2019	Kural Editörünün yazılması	44	8	-	-	0%										
İş Akış Motoru	30/05/2020	Arayüzün canlı kullanıma alınması	44	50	-	-	0%										

THANK YOU

Thank You

Contributions

Furkan METİN	Head of Department
Hasan TEKGÜL	Project Manager
Tuğrul GÜMÜŞKAYA	SQL Developer
İsmail AKGÜL	SQL Developer
Çetin KAYA	Analyst
Tuğba TUĞCU	Analyst
Cihan BARAN	SQL Developer
Yusuf UÇAN	SQL Developer
Yavuz ÇOBAN	SQL Developer
Özgür Volkan Suer	SQL Developer
Müge Naz Yıldırım	Analyst
Nesibe Gül	Analyst
Kudret Elif Berkman	SQL Developer
Ali Osman Bilgin	SQL Developer
Ümit Kaya	SQL Developer
Ezgi Sena Ertürk	Software Developer
Elif Şişman	Software Developer
İsmail Arslan	Analyst Intern