İzmir İktisat Dergisi
İzmir Journal of Economics

E-ISSN: 1308-8505 **Year:** 2023 **Vol:** 38 **No:** 1 **Pages:** 142-157
Recieved: 31.03.2022 **Accepted:** 04.08.2022 **Published Online:** 05.03.2023 **Doi:** 10.24988/ije.1096088
RESEARCH ARTICLE

Turkey's Labor Market During Covid-19 Pandemic: A K-Modes Analysis

Bige KÜÇÜKEFE [1] , Nilüfer KAYA KANLI [2]

Abstract

The COVID-19 pandemic has caused significant economic contractions and employment vulnerabilities for the economies, including Turkey. The pandemic exacerbated structural challenges related to high unemployment, low labor force participation, and widespread informality. This study aims to analyze the differences in the labor market between the 2019 and 2020 years in Turkey. For this purpose, we used the clustering method. While applying the clustering method, we used education type, gender, and age group data. Moreover, the study also employed information from employed, unemployed, and not in labor force data. We implemented a Machine Learning method, K-modes analysis, using the Turkish Statistical Institute's employment statistics and Labor Force Statistics Micro Datasets for 2019 and 2020.
Keywords: *Covid-19, Pandemic, Turkey, Labour Market, Employment, K-Modes.*
Jel Codes: *J21, J82*

Covid-19 Pandemisinde Türkiye İşgücü Piyasası: Bir K-Modes Analizi

Özet

COVID-19 küresel salgını, Türkiye de dahil olmak üzere ekonomilerde büyük daralmalar ve istihdam kırılganlıklarına neden oldu. Yüksek işsizlik, düşük işgücü katılımı ve geniş çaplı kayıtsız işgücüne bağlı yapısal problemleri şiddetlendirdi. Bu çalışmanın amacı, Türkiye'de işgücü piyasasındaki 2019 ve 2020 yıllarındaki değişimi incelemektir. Bu amaçla, kümeleme yöntemini kullandık. Kümeleme yöntemini uygularken, çalışan, işsiz ve iş gücüne dahil olmayanlar; eğitim durumu, cinsiyet ve yaş grubu verilerini kullandık. Bir makina öğrenmesi yöntemi olan K-Modes analizini, 2019 ve 2020 yılları için Türkiye İstatistik Kurumu'nun istihdam istatistikleri ve İşgücü İstatistikleri Mikro Veri Setleri'ne uyguladık.
Anahtar kelimeler: *Covid-19, Küresel Salgın, pandemi, İşgücü Piyasası, İstihdam, K-Modes.*
Jel Kodu: *J21, J82*

CITE (APA): Küçükefe, B., Kaya Kanlı, N. (2023). Turkey's Labour Market During Covid-19 Pandemic: A K-Modes Analysis. *İzmir İktisat Dergisi*. 38(1). 142-157. Doi: 10.24988/ije.1096088

[1] Dr., Tekirdag Namik Kemal University, Marmaraereglisi Vocational School, Tekirdag, Turkey,

EMAIL: bkucukefe@nku.edu.tr **ORCID:** 0000-0003-1945-3037

[2] Dr., Independent Researcher, Istanbul, Turkey , **EMAIL:** nilukaya@gmail.com **ORCID:**0000-0001-9049-3862

1. INTRODUCTION

The disruption caused by COVID-19 is not limited to individual health due to its ability to spread fast and cause death. The economic consequences of the pandemic are also destructive. The world economic output contracted by 3.5% in 2020 (IMF, 2021), leading to vast employment and working-hour losses. The global employment-to-population ratio decreased from 57.9 in 2019 to 54.6 in 2020. The groups most experienced a more significant decline are women, youth, and the medium and low-skilled workers in 2021 (ILO, 2021 a). Globally 8.8% of working hours were lost in 2020 compared to the fourth quarter of 2019. This loss is equivalent to 255 million full-time jobs and four times greater than during the global financial crisis in 2009 (ILO, 2021 b). According to Furman (2020), the synchronized shutdown in the first half of 2020 was the largest, fastest, and most comprehensive reduction in economic activity ever witnessed in the modern world. It affected supply and demand and was costliest for less advantaged households, whose members were the most likely to lose their jobs.

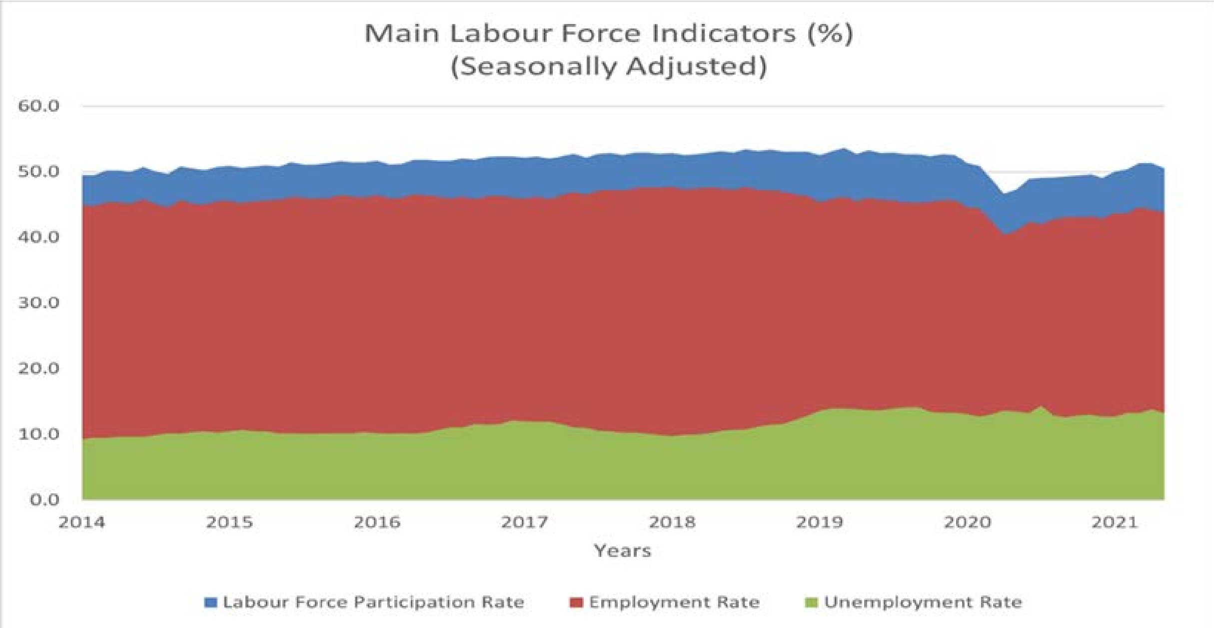
The COVID-19 pandemic has caused significant economic contractions for all economies, including Turkey, albeit to varying degrees, which implied a large shock on the employment market. In this period, people in vulnerable employment, especially in sectors that are more likely to be affected by the epidemic and for employees who cannot efficiently complete their work from home, are at risk of losing their jobs and livelihoods due to social distance and/or company closures. According to the "COVID-19 Business Impact and Needs Survey", four in every five SMEs (Small and Medium Sized Enterprises) were significantly affected by the COVID-19 crisis, while only 8% of all businesses stated that their businesses continue their routine activities in Turkey. The survey implied that, while 54% of enterprises reported a loss of workforce in the three big cities, 64% did so in other cities. 63% of businesses employing female workers, who make up more than half of their total workforce, reported job losses (Business for Goals Platform, 2020).

The problems of the Turkish labor market were complex even before the epidemic in Turkey. The COVID-19 shock worsened the already weakened labor market situation. Despite impressive economic performance since 2001, growth has primarily depended on foreign-currency credit booms and private sector debt since the 2008-2009 Global Financial Crisis (Şeker, Özen and Erdoğan, 2020). After the 2001 crisis, the unemployment problem could not be solved despite the high growth rates with the IMF-supported program. While the unemployment rate was 6.5% in 2000, the lowest level in the last 40 years, it increased to 10.84% in 2004. In 2009, because of the banking crisis that started in the USA, taking a global form, and the crisis spreading to the real sector, the unemployment rate rose to another peak of 12.55. Afterward, it started a downward trend and fell to 8.15 in 2012. While the economic growth fluctuated between 8.48% and 0.89% in the 2013 – 2019 period, the unemployment rate increased from 8.73% to 13.67%, and it was argued by many researchers that Turkey grew without employment and was caught in the middle-income trap (Development Ministery, Özel İhtisas Komisyonu Raporu, 2018; Baştav, (2021); Uğurluklukol and Tükenmez, 2019, Worldbank, 2021).

Unemployment fluctuations are greatly affected by changes in supply and demand. The pandemic exacerbated structural challenges related to high unemployment, low labor force participation, and widespread informality.

Figure 1 indicates that the unemployment rate has increased while the labor force participation and employment rates were decreasing, especially during the first quarter of 2020 (TUIK, 2020).

Figure 1: Labor force participation rate, employment rate, the unemployment rate



Source: TurkStat

This study aims to analyze the difference in the labor market between 2019 and 2020 in Turkey. We are implementing K-modes analysis using the Turkish Statistical Institute's employment and labor force micro datasets (TurkStat, 2019, 2020).

1.1 Motivation

The pandemic worsened the structural challenges of high unemployment and low labor force participation. This study aims to analyze the difference in the labor market between 2019 and 2020 in Turkey. For this purpose, we used the clustering method. While applying the clustering method, education type, gender, and age group data were used. Moreover, employed, unemployed, and not in labor force data were employed. We implemented a Machine Learning method, K-modes analysis, using the Turkish Statistical Institute's employment statistics and Labor Force Statistics Micro Datasets for 2019 and 2020.

1.2 Contribution

We implement the K-modes clustering algorithm, well-known for its clustering efficiency and stability. Our first contribution to the literature is using the K-Modes algorithm to analyze the labor market with the data on labor force status, gender, education type, and age group of household members. Secondly, we compared the clusters of 2019 and 2020 to observe the changes in the labor market in 2020, when the significant changes were driven by the COVID-19 pandemic.

2. RELATED WORKS

There are numerous studies using socioeconomic microdata for clustering. All these studies chose suitable cluster methods for different data types and purposes. We review these studies in the literature section to understand how employment and other socioeconomic variables are used in the cluster analysis. For example, Özdemir and Demir (2019), classified the socioeconomic status of individuals in Turkey using SILC (Survey of Income and Living Conditions) data for 2015. They applied latent class (LCA) analysis, K-modes clustering, and random forest analysis to define individuals' socioeconomic profiles. For this purpose, Özdemir and Demir (2019) handled 36.036 cross-sectional data of individuals and determined variables affecting the individuals' income. Ten clusters were obtained in cluster analysis, and education, occupation, and age variables were more important than other variables used in this research for the random forest analysis.

D'Urso and Massari (2019) proposed a fuzzy clustering model with mixed features and conducted two empirical applications to show the effectiveness of the proposed clustering algorithm. Their second empirical application aimed to cluster individuals based on attributes of different types, i.e., socio-demographic and news consumption habits. Then, the study investigated the attitudes of individuals in each cluster towards various issues related to migration, economic conditions, and European Union integration. For this study, data is selected for Italy drawn from the sixth wave of the survey, held in 2012, and variables were grouped into numeric, interval-valued, and categorical.

Grane, Albarran, and Lumely (2020), visualized profiles of aged European individuals to understand the risk to their health and social wellbeing. This study's data was taken from the Survey of Health, Ageing and Retirement in Europe (SHARE) for 18 countries and 30 mixed-type variables. These variables were related to socio-demographic information, such as; gender, age, education, employment (etc.), and were used for implementing the k-prototypes clustering algorithm. K-prototypes used the K-modes algorithm for mixed-type variables. When the cluster was created, they made a profile of each cluster's average number. Thus, the profiles have been ranked by taking the mean of the mean values. The profiles with higher mean values in the indices were more disadvantaged group than those with higher indices. This study is very useful for cluster analysis. Variables such as gender, age, education, and employment are also used in our analysis. Making profiles of the average number of each cluster will also be considered further.

Brada, Gajewski, and Kutan (2021) examined resiliency and the ability in Central and Eastern Europe (CEE) following the 2008 Global Financial Crisis in employment and found significant clusters of high-performing and low-performing areas. Using these analyses, they estimated the coefficient from their recovery equation and stimulated the impact of COVID-19 on regions. The paper showed that recovery from the effects of the pandemic on employment would be slow.

For Thailand, with the K-means clustering method, Wasi et al. (2020) illustrated that more than half of registered workers left the formal sector either seasonally or permanently long before their retirement age. The employment data from Thailand Social Security records included millions of individuals' work history information, such as job tenure, the number of times an employee exited and returned to the formal sector, and out-of-formal sector duration. As the data are numerical, K-means clustering can be used effectively in this study.

Macroeconomic indicators were also used in the existing literature on clustering. As an example, Kutlar, Gülmez, and Oncel (2021) examined OECD countries, including Turkey, whether the clustered countries showed similarity in terms of economic impact before and after the pandemic. The analysis used multidimensional scaling and cluster methodology with export, import, inflation, unemployment, private consumption, and growth rate variables.

Some articles evaluate the impact of COVID-19 on employment from other perspectives. Kalenkoski and Pabilonia (2020) examined the initial impact of COVID-19, shutdowns on employment, and hours of unincorporated self-employed workers regarding gender, couple status, and parental status. The random effect and difference-in-differences models were used to examine the initial differential impacts of the COVID-19 pandemic shutdowns. This study shows the adverse effects of being a coupled woman and being a self-employed woman on retention in employment. Şeker, Özen and Erdoğan (2020) analyzed employment vulnerability in different sectors due to COVID-19 in Turkey and proposed the "Employment Vulnerability Index" which implies working from home index. As a result, they found around 7 million workers are at the risk of losing their jobs due to the economic impacts of COVID-19. Bauer and Weber (2020) evaluated how much the containment measures affect unemployment in Germany. Li, Song and Wu (2020) analyzed the deterioration of households' liquidity caused by unemployment. Moen, Pedtke and Flood (2020) examined intersectional COVID-19 Employment Effects by Age, Gender, Education, and Race/Ethnicity. Coibion, Gorodnichenko and Weber (2020) estimated job losses in the COVID-19 pandemic. Cajner et al. (2020) documented that businesses cut nominal wages for nearly 7 million workers while abandoning regularly scheduled wage increases for others. Finally, Alon, Coskun, Doepke, Koll, and Tertilt (2021) showed women's versus men's employment in regular and pandemic recessions across industries and occupations.

3. METHOD

We used the Labor Force Micro Dataset (LFMD) prepared by TurkStat (Turkish Statistical Institute, 2019, 2020). Data provided in CSV format and Python 3.0 is used for data mining and clustering analysis. For Household Labour Force Survey (2019, 2020), a two-stage stratified method was used. Based on the address, a rotation pattern is formed to ensure a 50% overlap between two consecutive periods and the same periods of the two successive years. Eight subsamples were used in each period. In the study design, the sample size has been equally distributed to the weeks, which will be applied in each term. In this part, we briefly give information about labor force status, gender, education type, and age group of household members. The percentages of the variables are our calculations.

The LFMD (TurkStat 2019,2020) variables and the codes that we used in the K-modes analysis are as follows:

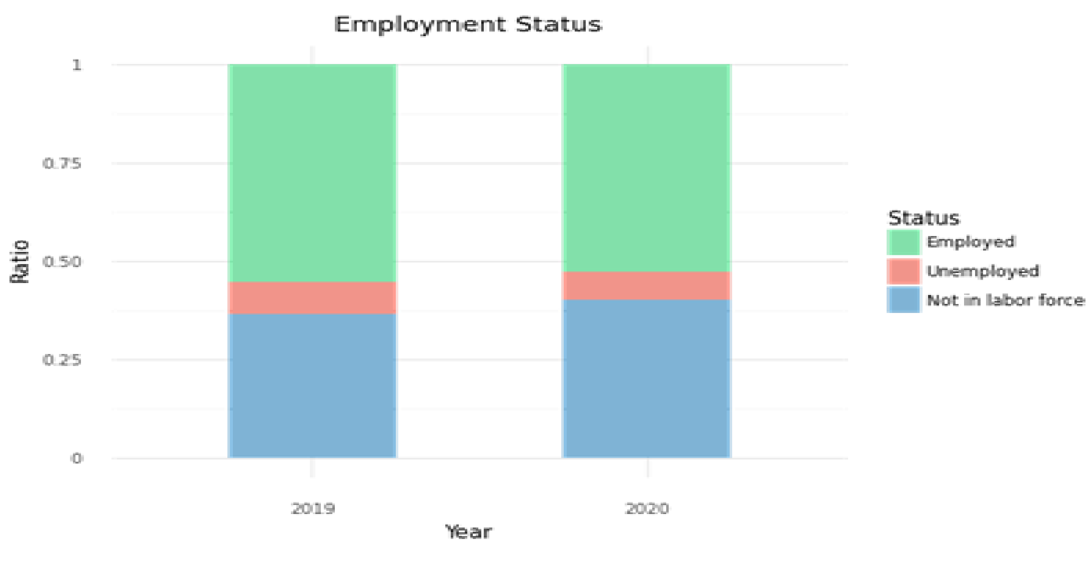
3.1 Variables

3.1.1 Labor force status

Labor force status of household members are coded as: 1. Employed, 2. Unemployed, 3. Not in labor force in LFMD (TurkStat 2019,2020).

Figure 2 and Table 1 depicts, while persons who do not participate in the labor force had an increasing trend, the number of employed and unemployed decreased according to the LFMD. Labor force status of household member are coded as: 1. Employed, 2. Unemployed, 3. Not in labor force in HLFS (TurkStat 2019,2020).

Figure 2: Labor force Status



Source: TurkStat

Table 1: Labor Force Status

Labor Status	Year	Frequency	% Percentage
Employed	2019	139130	0.551270
Unemployed	2019	20831	0.082538
Not in Labor Force	2019	92420	0.366192
Employed	2020	168293	0.525059
Unemployed	2020	23557	0.073496
Not in Labor Force	2020	128672	0.401445

Source: TurkStat

3.1.2 Gender

Gender of household members are coded as: Gender: Male (1), Female (2) in HLFS (TurkStat 2019,2020).

Table 2 and Table 3 indicate frequency and percentage of women and men employed, unemployed and not in the labor force. A deeper look to Table 1 shows, while the rate of female employees was only 35.85% in 2019, this number decreased further to 33.34% in 2020. It is interesting to note that while the number of unemployed women increased by 305, the percentage of unemployment decreased from 6.68 % to 5.46 %. From 2019 to 2021, 25000 women gave up looking for a job for various reasons and were included in the not in labor force part. For 2020, 61.19% of women were not in the labor force, which is a very high rate. From 2019 to 2020, the number of female employees increased by 8261.

Table 2: Female Labor Status

Female Labor Status	*Year*	*Frequency*	*Percentage (%)*
Employed	*2019*	*46490*	*35.8586*
Unemployed	*2019*	*8663*	*6.6819*
Not in Labor Force	*2019*	*74495*	*57.4594*
Employed	*2020*	*54751*	*33.3439*
Unemployed	*2020*	*8968*	*5.4616*
Not in Labor Force	*2020*	*100482*	*61.1945*

Source: TurkStat and Authors' calculations

The situation is quite the same on the male employment trends side where employment and unemployment percentage decreased, but not in labor force percentage increased. The significant difference between female and male employment is the size of the employment status.

Table 3: Male Labor Status

Male Labor Status	Year	Frequency	Percentage (%)
Employed	2019	92640	75.4809
Unemployed	2019	12168	9.9142
Not in Labor Force	2019	17925	14.6049
Employed	2020	113542	72.6339
Unemployed	2020	14589	9.3327
Not in Labor Force	2020	28190	18.0334

Source: TurkStat and Authors' calculations

3.1.3 Education:

The survey question and Education codes in HLFS are as follows:

What is the latest educational institution/level you graduated from?

0. Literate but not completed any educational institution

1. Primary school (4 or 5 years)

2. Lower secondary, Vocational, and technical secondary school or Primary education

3.1. Upper secondary school (High school)

3.2. Vocational and technical high school

4. 2 or 3 years higher education or faculty or 4 years higher education or faculty

5. Masters's degree (5 or 6 years faculty included) or Doctorate

Table 4 indicates the frequency and percentage of an individual's education level.

Table 4: Education

Education	Year	Freq.	Percentage (%)	Year	Freq.	Percentage (%)
0	2019	1792	10.0251	2020	2586	0.91735
1	2019	6195	34.5607	2020	9316	33.0472
2	2019	2592	14.4603	2020	4462	15.8283
3.1	2019	3114	17.3724	2020	4814	17.0770
3.2	2019	1882	10.4993	2020	3080	10.9259
4	2019	2206	12.3068	2020	3704	13.1394
5	2019	139	0.7755	2020	228	0.8088

Source: TurkStat and Authors' calculations

3.1.4 Age Groups

Age groups are coded as: Group of completed age: 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59

Table 5: Age Groups

Age Group	Year	Freq.	Percentage (%)	Year	Freq.	Percentage (%)
20-29	2019	21385	23.1389	2020	30717	23.8723
30-39	2019	20685	22.3794	2020	28134	21.8649
40-49	2019	21096	22.8262	2020	29730	23.1053
50-59	2019	29256	31.6555	2020	40091	31.1575

Source: TurkStat and Authors' calculations

The number of individuals in LFMD 2019 and 2020 are 239.050 and 474.514, respectively. We remove individuals in groups '0-19', '60+' and set the age group intervals to 10 instead of 5, to make it more suitable for the analysis and to make comparisons easily. The number of total individuals used in K-modes analysis remained at 572.903 for given years.

Clustering is a crucial research tool to discover underlying structures in unlabeled data and is used in pattern recognition, data mining, statistics, and machine learning applications. In clustering analysis, the K-Means algorithm is commonly and efficiently used for large data sets. It has been applied to many areas, from machine learning (unsupervised learning) to computer graphics (Arthur and Vassilvitskii, 2006). K-Means method (Anderberg, 1973), groups the numerical data according to existing similarities among them in k clusters where Euclidian distance is used as a distance measure. However, working on only numeric values is insufficient for real-world clustering data. Data mining applications require the processing and exploration of numeric, categorical, or both types of data (Khan and Ahmad, 2013). Examples of categorical variables in our research are {male, female}and {employed, unemployed, not in labor force}. Categorical data cannot be ordered. Thus, it is not possible to find Euclidian distance between them. Huang (1997) first introduced K-Modes as a clustering method that can analyze numerical and categorical data. K-Modes algorithm measures dissimilarity with a frequency-based method and replaces means with modes.

3.2 K-Modes Algorithm

K-Modes algorithm extends the K-Means process to cluster categorical data and uses a frequency-based method. This new dissimilarity measure replaces means of clusters with cluster centers (modes) to minimize the cost function. Huang (1998) also shows that the K-Modes algorithm is faster than K-Means as it needs fewer iterations for convergence.

The new dissimilarity measure (Hamming distance) can be described. Let X and Y be two categorical data objects in n categorical variables. The measured difference (X,Y) between X and Y can be described by the total mismatches of the corresponding feature categories of the two objects. Smaller the number of mismatches shows more similarity between two objects. The distance function is calculated as follows:

$$d(X,Y) = \sum_{i=1}^{n} \delta(x_i, y_i) \tag{1}$$

where $\delta(x_i, y_i) = \begin{cases} 0 & (x_i = y_i) \\ 1 & (x_i \neq y_i) \end{cases}$

Let K be a set of k categorical objects in n categorical attributes, $N_1, N_2, N_3, \ldots, N_n$.

where Eq. (1) distance function, the cost function becomes:

$$C(Q) = \sum_{j=1}^{m} d(M_j, Q_j) \tag{2}$$

where M_j: j th element , Q_j: Nearest cluster center of M_j

The K-modes algorithm minimizes the cost function defined in Eq. (2). K-modes algorithm follows these steps (Huang, 1997) :

Randomly select k starting point (mode), one for each cluster as the initial cluster centers.

Calculate the dissimilarities and allocate an object to the cluster with its mode closest to d. Update the mode of the cluster after each allocation.

After all the objects are divided into clusters, retest the dissimilarity of objects according to the current model. If an object is found, the closest mode belongs to another cluster rather than its cluster. The current one assigns the object to that cluster and updates the modes of both sets.

Repeat step 3 until there is no re-allocation required.

Determining the optimal number of clusters (k) in a k-modes clustering problem is essential. One useful method for obtaining the optimal value of k is the elbow method which plots various values of cost with changing k. The scree plot from the elbow method is provided in Fig.2 and Fig.3 for the years 2019 and 2020, respectively. The elbow suggests the optimal value for k is 5 and 10 for both years. We run the developed program and choose 10 as an optimal k value.

Figure 3a: Scree Plot, 2019

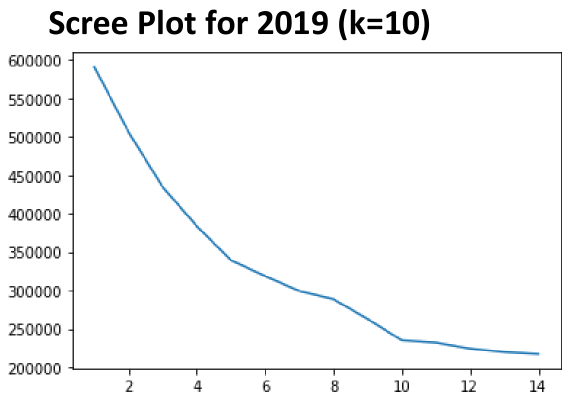
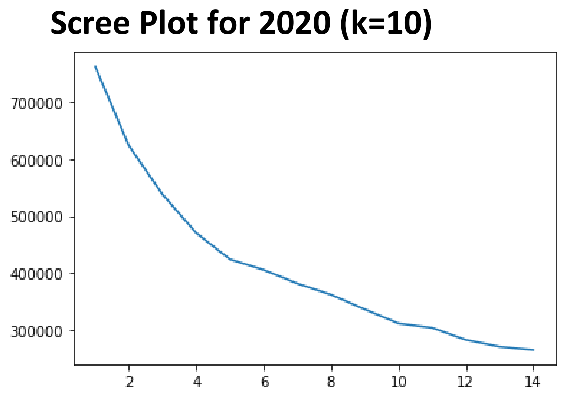


Figure 3b: Scree Plot, 2020



Source: Authors' calculations

4. RESULTS

Table 6 and Table 7 show our clusters applied to the 2019 and 2020 LFM Dataset, respectively. For 2019, Cluster 0 consists of a female with primary school education, employed, and the 20-39 age group. Cluster 1 contains male, University educated, not in the labor force 40-49 age group. Cluster 2 includes male, the lower secondary school educated, employed, and 20-29 age group. Female, the lower secondary school, educated not in labor force and 50-59 age group are in 3rd cluster. Finally, last cluster 4 includes male primary school educated, employed, and 50-59 age group.

2020 Survey clusters are very similar to 2019 Survey clusters. In cluster 0 age group changed 30-39 to 40-49 and in cluster 1 age group changed 40-49 to 30-39.

Table 6: Cluster Types (2019 Survey)

CLUSTERS	GENDER		CLUSTERS EDUCATION		EMPLOYMENT STATUS		AGES
	Codes	Gender	Codes	Education	Codes	Status	
0	2	Female	1	Primary school	1	Employed	30-39
1	1	Male	4	University	3	Not in labor force	40-49
2	1	Male	2	Lower secondary school	1	Employed	20-29
3	2	Female	2	Lower secondary school	3	Not in labor force	50-59
4	1	Male	1	Primary school	1	Employed	50-59

Source: TurkStat and Authors' calculations

Table 7: Cluster Types (2020 Survey)

CLUSTERS	GENDER		CLUSTERS EDUCATION		EMPLOYMENT STATUS		AGES
	Codes	**Gender**	**Codes**	**Education**	**Codes**	**Status**	
0	2	Female	1	Primary school	1	Employed	40-49
1	1	Male	4	University	3	Not in labor force	30-39
2	1	Male	2	Lower secondary school	1	Employed	20-29
3	2	Female	2	Lower secondary school	3	Not in labor force	50-59
4	1	Male	1	Primary school	1	Employed	50-59

Source: Authors' calculations

Cluster parts and sizes are as follows:

4.1 Cluster Parts

4.1.1 Labor force status

Table 8 shows K-modes cluster sizes according to the labor force status, divided into three parts: employed, unemployed and not in the labor force. Cluster 0 is the largest cluster, comprising 40% of respondents in the 2019 Survey. At the same time, employees constitute the majority of this cluster with a rate of 64%, and not in the labor force takes second place with 26%. The other four clusters are approximately the same as each other in size. Clusters 1,2,3,4 are composed of 32.81%, 84,76%, 0.85%, %82.56 employed, % 12.91, %12.81, %1.08, %4.29 unemployed and 54.28%, 4.24%, 98.07%, 13.15% not in labor force, respectively in 2019 survey.

Cluster 0 is the largest cluster also in the 2020 survey. The feature of cluster 0 is that it mainly consists of females, primary school graduates, employees, and in the 40-49 age group. In 2020, when COVID was effective, the rate of the labor force remained approximately the same, while the rate of working people increased from 64.76 to 65%, and the unemployment rate decreased from 9.01 to 7.60% in cluster 0.

Cluster 1 consists mainly of men, university graduates in the 30-39 age group, and those not in the labor force. The employment level of this cluster, which consists of university educated, aged 30-39 men who are not in the workforce, increased from 32.81% to 34.46%, while unemployment decreased from 12.91% to 10.71%, and the not in labor force rate remained the same.

In Cluster 2, there are mostly lower secondary school graduates, 20-29 aged, employed men. In this cluster, while the employed rate decreased slightly from 84.76% to 84.70%, the unemployment rate decreased from 12.81% to 11.53%, and not in labor force rate increased from 2.42% to 3.76% in 2020.

Cluster 3 comprises mostly female, lower secondary school graduates aged 50-59 and not in the labor force. In this cluster, the rates didn't change too much. In cluster 3, employed, unemployed, and not in labor force rates are 0.80%, 1.15% and 98.05%, respectively in 2020.

In table 5 and 6, male, primary school graduates, employed and aged 50-59 constitute our last cluster predominantly. In cluster 4, the employment rate decreased drastically from 82.56% to 76.53% in 2020. The unemployment rate increased slightly from 4.29% to 4.60%, and not in labor force increased gradually from 13.15% to 18.87%.

Table 8: Cluster Sizes (2019 and 2020 Survey)

	2019 CLUSTERS					2020 CLUSTERS				
	0	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
Employed	66732	12278	31997	346	27777	82541	21239	38491	436	25586
Unemployed	9283	4832	4836	436	1444	9552	6600	5241	626	1538
Not in Labor Force	27031	20313	915	39738	4423	33580	33802	1709	53271	6310
Total	103046	37423	37748	40520	33644	125673	61641	45441	54333	33434

Source: Authors' calculations

Figure 4a: Scree Plot, 2019

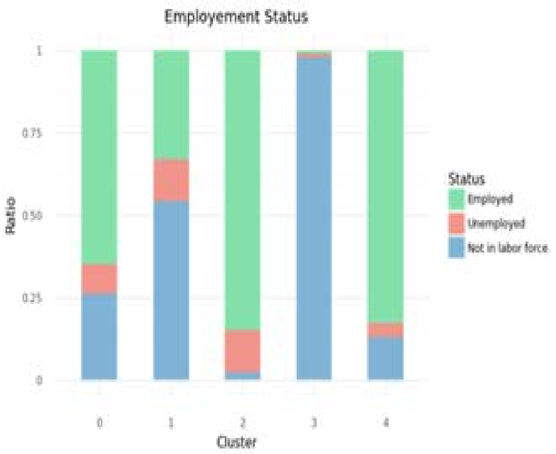
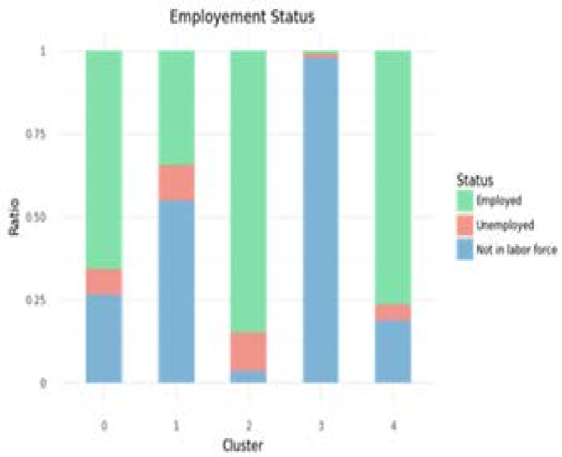


Figure 4b: Scree Plot, 2020



Source: Authors' calculations

4.1.2 Age Group

Table 9 shows age group cluster sizes:

Table 9: Age Group Cluster Sizes (2019 and 2020 Survey)

	2019 CLUSTERS					2020 CLUSTERS				
	0	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
20-29	11911	8060	25849	13446	-	13640	12695	30949	17479	-
30-39	57359	1784	6047	2995	-	25207	41826	6770	4810	6129
40-49	25764	24522	3335	1578	11723	77420	2549	4606	2003	-
50-59	8012	3057	2517	22501	21921	9406	4571	3116	3041	27305

Source: Authors' calculations

Figure 5a: Age Group in Clusters, 2019

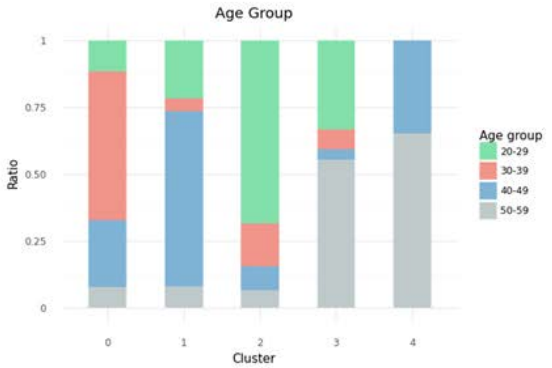
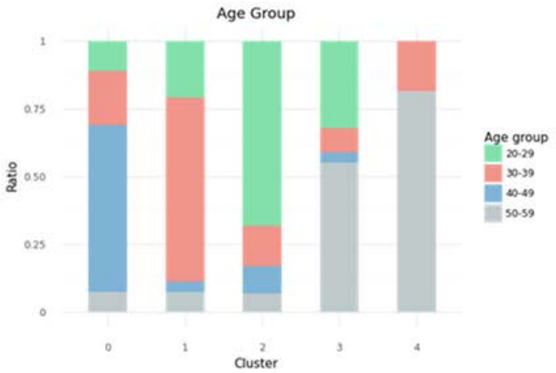


Figure 5b: Age Group in Clusters, 2020



Source: Authors' calculations

Table 9 shows K-modes cluster sizes according to age group, which is divided into four parts as; 20-29,30-39,40-49 and 50-59. 56% of the respondents in Cluster 0, the largest cluster, are in the age group of 30-39 in the 2019 survey. Clusters 1,2,3 are composed of 22%, 68%, and 33% in the age group of 20-29, while Cluster 4 doesn't have any 20-29 group respondent; 5%,16%, and 7% at the 30-39 age group and Cluster 4 again don't have any respondents in this age group. At Clusters 1,2,3,4. 66%, 9%, 4% and %35 of respondents are at 40-49 age group and 8%, 7%, 56% and 65% are at 50-59 age group.

The feature of cluster 0 is that it mainly consists of the 40-49 age group in 2020 when COVID-19 was effective. In 2020, the rate of the 40-49 age group increased from 25% to 62%, and the 30-39 age group decreased from 56% to 20% in Cluster 0.

4.1.3 Education

Table 10 shows education cluster sizes:

Table 10: Education Cluster Sizes(2019 and 2020 Survey

	2019 CLUSTERS					**2020 CLUSTERS**				
	0	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
0	10162	5885	956	8382	847	10785	7957	934	9617	910
1	41071	2630	306	10967	26917	57426	1850	339	15060	25139
2	2812	967	24006	12692	2385	3190	2155	28640	17566	-
3.1	11181	7930	2960	4470	1758	12354	12942	3527	6231	2056
3.2	10137	6304	4065	2959	1390	11054	10447	5273	4460	1687
4	24717	12741	5136	921	2385	27180	24409	6298	1219	3139
5	2966	966	319	129	347	3684	1881	430	180	503

Source: Authors' calculations

Figure 6a: Education Group in Clusters, 2019

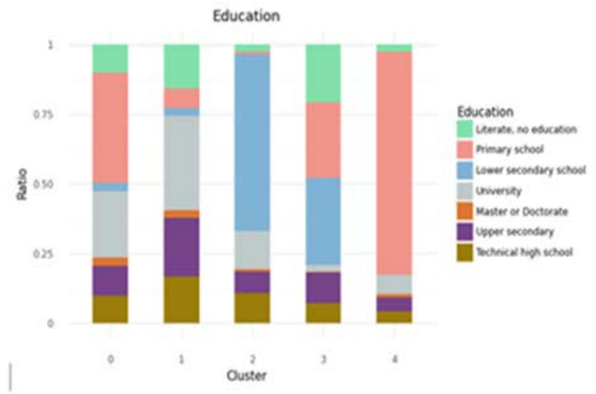
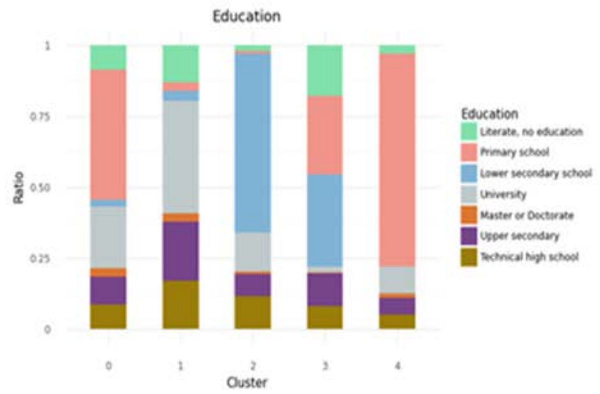


Figure 6b: Education Group in Clusters, 2020



Source: Authors' calculations

Table 10 shows K-modes cluster sizes according to education level, which is divided into seven parts as (0) literate but not completed any educational institution, (1) primary school, (2) Lower Secondary, Vocational and technical secondary school or Primary education , (3.1) Upper secondary school (High school), (3.2) Vocational and technical high school, (4) 2 or 3 years higher education or faculty or 4 years higher education or faculty and finally (5) Master degree (5 or 6 years faculty included) or Doctorate

Clusters 0,1,2,3,4 are composed of 40%,7%, 1%, 27% 75% primary school graduates while 24%,34%, 14%, 2% 7% for 2 or 3 year higher education or faculty or 4 years higher education; In 2020, at Clusters 0,1,2,3,4. 46%, 3%, 1% and 28% and 75% were primary school graduates and 22%, 40%, 14% and 2% and 9% were higher education or faculty or 4 years higher education graduates.

5. CONCLUSION

The COVID-19 pandemic has caused many economic and social destructive consequences. The labor market affects the economy in many ways, and the labor market is among the crucial factors that determine the state of the economy. In this respect, microdata is a valuable resource for labor market research. However, there are not many studies using microdata in this area. The reason for this situation is the scarcity of methods that can use microdata. We think that state of art machine learning algorithms has offered new ways to use microdata in recent years. Primarily, K-modes extend the K-Means algorithm and uses numerical and categorical data, which is very useful for the LFMD. In this study, we used the K-Modes algorithm to analyze the labor market with the labor force status, gender, education type, and age group of household members data. Moreover, we compared the 2019 and 2020 years clusters. For comparing clusters making profiles of the average number of each cluster can also be considered in a further study.

Modern unemployment theory has come a long way. In modern economics, structural unemployment arises when there is a mismatch between the skills demanded and supplied in a given area or an imbalance between the supplies of and demands for workers across areas (Ehrenberg, Smith and Hallock, (2021)). The COVID-19 pandemic has affected both supply and demand sides of the economy. With the cluster analysis, we were able to see the changes in the demand for labor in Turkey as age, education, and gender.

To perform K-modes analysis, first, we determined the number of clusters. The Elbow method suggests from our data, that the optimal cluster number is 5. In our K-Modes analyses, we had 5 clusters in 2019 and 2020. For 2019, Cluster 0 consists of females with primary school education, employed, and the 20-39 age group. Cluster 1 contains males, and University education, not in labor force individuals within the 40-49 age group. Cluster 2 includes males that are lower secondary school educated, employed and the 20-29 age group. Female, lower secondary school educated not in labor force individuals that are in the 50-59 age group are in 3rd cluster. Finally, last cluster 4 includes males primary school educated, employed and in the 50-59 age group. Our cluster analysis results are as follows:

In 2019, the age group in Cluster 0 was 30-39, in 2020 it was 40-49. In Cluster 1, on the other hand, the age group has changed and vice versa. We can say that, in 2020, the age group increased in cluster 0, which consisted of mostly primary school graduates and female employees. The age group dropped to 30-39 in Cluster 1, which consisted mostly of male university graduates and those not in the workforce in 2020. Another result in cluster 0 is the unemployment rate decreased from 9.01 to 7.60%.

In Cluster 2, which consists of mostly lower secondary school graduates, 20-29 aged, employed men unemployment rate decreased from 12.81% to 11.53%.

Cluster 3 consists of mostly females, lower secondary school graduates aged 50-59 and not in labor force individuals. For this cluster the rates didn't change too much. There was an increase in the not in labor force rate in all clusters except the third cluster between 2019 and 2020. Meanwhile, in the third cluster not in the labor force remained approximately the same.

In cluster 4, which consists of mostly males, primary school graduates, employed and aged 50-59 individuals, the employment rate decreased from 82.56% to 76.53% in 2020.

REFERENCES

Alon, T., Coskun, S., Doepke, M., Koll, D., & Tertilt, M. (2021). From Mancession to Secession: Women's Employment in Regular and Pandemic Recessions, No. w28632. National Bureau of Economic Research.

Anderberg, M.R. (1973), Cluster analysis for applications, Academic Press, New York.

Arthur D. & Vassilvitskii, S. (2006). How Slow Is The K-Means Method?, In Proceedings of the 22. Annual Symposium on Computational Geometry,144-153. 22nd ACM Symposium on Computational Geometry (SoCG), Arizona. New York: Association for Computing Machinery.

Baştav, L. (2021). Turkish Economic Growth:(Non) Labour Creating? (2000-2018). Sosyoekonomi, 29(50), 35-50.

Bauer, A., & Weber, E. (2021). COVID-19: how much unemployment was caused by the shutdown in Germany?, Applied Economics Letters, 28(12), 1053-1058.

Brada, J. C., Gajewski, P., & Kutan, A. M. (2021). Economic resiliency and recovery, lessons from the financial crisis for the COVID-19 pandemic: A regional perspective from Central and Eastern Europe. International Review of Financial Analysis, 74(101658).

Business for Goals Platform (2020). "Survey on Impact of COVID-19 on Enterprises and Needs" https://www.tr.undp.org/content/turkey/en/home/presscenter/pressreleases/2020/04/B4G-anket.html

Cajner, T., Crane, L.D., Decker, R. A., Grigsby, J., Hamins-Puertolas, A., Hurst, E., Kurz, C., & Yildirmaz, A. (2020). The US Labor Market During The Beginning Of The Pandemic Recession, BFI Working Paper.

Coibion, O., Gorodnichenko, Y., & Weber, M. (2020). Labor markets during the COVID-19 crisis: A preliminary view, No. w27017. National Bureau of Economic Research.

Development Ministery, (2018). On Birinci Kalkınma Planı (2019-2023), İşgücü Piyasası ve Genç İstihdamı Özel İhtisas Komisyonu Raporu.

D'Urso, P. & Massari, R. (2019). Information Sciences, 505, 513-534.

Ehrenberg, R. G., Smith, R. S., & Hallock, K. F. (2021). Modern labor economics: Theory and public policy. Routledge.

Furman, J. (2020). Protecting people now, helping the economy rebound later, VoxEU.org, 18 March

Grane, A., Albarran, I., & Lumely R. (2020). International Journal of Environmental Research and Public Health, 17(7747).

Hall, R. E. (2003). Modern theory of unemployment fluctuations: Empirics and policy applications. American Economic Review, 93(2), 145-150.

Huang, Z. (1997). A fast-clustering algorithm to cluster very large categorical datasets in data mining. In Research issues on data mining and knowledge discovery. Data Mining and Knowledge Discovery, 3(8), (34-39)

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery, vol 2, 283–304. https://doi.org/10.1023/A:1009769707641

ILO (2021). ILO Monitor: COVID-19 and the world of work. Seventh edition. Updated estimates and analysis.

ILO (2021). ILO Monitor: COVID-19 and the world of work. Eighth edition. Updated estimates and analysis

International Monetary Fund (2021). World Economic Outlook Update, July 2021: Fault LinesWidenintheGlobalRecovery.https://www.imf.org/en/Publications/WEO/Issues/2021/07/27/world-economic-outlook-update-july-2021, (accessed Aug 14, 2021).

Li, J., Song, Q., Peng, C. & Wu, Y. (2020) COVID-19 Pandemic and Household Liquidity Constraints: Evidence from Micro Data, Emerging Markets Finance and Trade, 56(15), 3626-3634, DOI: 10.1080/1540496X.2020.1854721

Kalenkoski, C. M., & Pabilonia, S. W. (2020). Initial Impact of the COVID-19 Pandemic on the Employment and Hours of Self-Employed Coupled and Single Workers by Gender and Parental Status. IZA Discussion Paper, No 13443.

Kutlar, A., Gülmez, A., & Öncel, A. (2021). The Analysis of the Effect of Covid 19 on Macroeconomic Indicators via MDS and Clustering Methods.

Moen, P., Pedtke, J. H., & Flood, S. (2020). Disparate disruptions: Intersectional COVID-19 employment effects by age, gender, education, and race/ethnicity. Work, Aging and Retirement, 6 (4), 207-228.

OECD, (2021), OECD Economic Surveys: Turkey 2021.

Ozdemir, O. & Demir, I. (2019). International Journal of Sciences: Basic and Applied Research (IJSBAR),48(7), 110-138

Shehroz S. Khan, & Amir Ahmad, Cluster center initialization algorithm for K-modes clustering, Expert Systems with Applications, 40(18), 2013, 7444-7456,

Şeker, S. D., Özen, E.N. & Erdoğan, A.E. (2020). Jobs at Risk in Turkey: Identifying the Impact of COVID-19. Social Protection and Jobs Discussion Paper;No. 2004. World Bank, Washington, DC. World Bank. https://openknowledge.worldbank.org/handle/10986/34360 License: CC BY 3.0 IGO. (accessed Sep. 23, 2021).

Uğurlukol, E., & Tükenmez, N. M. (2019). Middle income trap and Turkey: Ways of getting out. BRAND. Broad Research in Accounting, Negotiation, and Distribution, 10(1), 37-51.

Wasi, N., Ayudhya, C. D. N., Treeratpituk, P., & Nittayo, C. (2021). Understanding a Less Developed Labor Market through the Lens of Social Security Data, No. 147. Puey Ungphakorn Institute for Economic Research.

Worldbank (2021a). Gross Domestic Product Growth (GDP) (annual %) – Turkey, https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=TR, (accessed July 31, 2021).

Worldbank (2021b). Unemployment, total (% of total labor force) (modeled ILO estimate) - Turkey https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=TR, (accessed July 27, 2021).

TUIK (2019). Labor Force Statistics Micro Data Set.

TUIK (2020). Labor Force Statistics Micro Data Set.

TUIK(2021). İşgücü İstatistikleri, https://data.tuik.gov.tr/Kategori/GetKategori?p=istihdam-issizlik-ve-ucret-108&dil=1, (accessed July 22, 2021).