Bilkent University

Department of Computer Engineering

Senior Design Project

Project name: DAOS (Dealership Assistance and Optimization System)

Final Report

Metehan Kesekler, Kerem Tuna Özlü, Mustafa Said Sarı, Sarp Tuğberk Topallar

Supervisor: Uğur Doğrusöz

Jury Members: M.Mustafa Özdal and Mehmet Koyutürk

Progress Report

May 3, 2018

This report is submitted to the Department of Computer Engineering of Bilkent University in partial fulfillment of the requirements of the Senior Design Project course CS491/2.

1.Introduction

When dealerships are given by companies to third parties, two of the most essential concerns are legality and profitability for such a dealership. Both of these issues can be addressed by selecting the appropriate location for the dealership. Finding such proper locations manually is very troublesome, because one must obtain information about the legal restrictions and regulations about the area. These restrictions and regulations may differ from province to province, even sometimes municipalities within the same province enforce different regulations. In addition to that, profitability potential is very hard to measure and may be affected by many factors such as weather, time of the day, seasonal activities of the locals, shopping demographic of the area etc. Even if these potential factors are successfully be identified initially, they may also change later on. Luckily, all these problems are dependent of the location of the dealership and using a geographic information system (GIS) can be helpful to solve all of them.

The issue of legality is very area sensitive and different constraints must be met for different types of businesses. For example, liquor and tobacco shops cannot be nearer than 100 meters to schools, dorms or sanctuaries by law. Therefore, having such sensitive places as points of interest (POI) in DAOS and display them on the actual map would certainly be very beneficial for companies trying to pick up the right venue for the regulatory assessment.

When the profit potential of a dealership is considered, the area can be manually inspected. But manual inspection can be made for a limited amount of time and therefore may not be enough to cover the overall trend for the area. Sending a team to different cities around the country for field surveys is also very costly. DAOS will present the data about spending potential of the area such as number of residents in the area and the foot traffic of the streets. Data provided by DAOS can be interpreted to become aware of the trends about the area even it could be used for making suggestions about possible business opportunities since the information about the area is accumulated in the database.

After the dealership becomes operational, its performance must be continuously observed. Because, the habits of the customers are not static and businesses are very sensitive to competition. So, a very profitable location for a dealership may not be still profitable so monitoring the operations of that dealership can be very important for making a decision about allocating more resources there or closing it entirely. Monitoring the sales of different branches only provides information about quantity, but considering branch sales

according to their local shopper profile will give their actual performance metric for evaluation.

Our solution, DAOS with its enhanced user interface, interactive and color coded maps and broad capabilities about different types of businesses, provides the answers for the following questions “Can this dealership be opened here?”, “Should this dealership be opened here?” and “Is this dealership still profitable?”

2.Final Architecture and Design

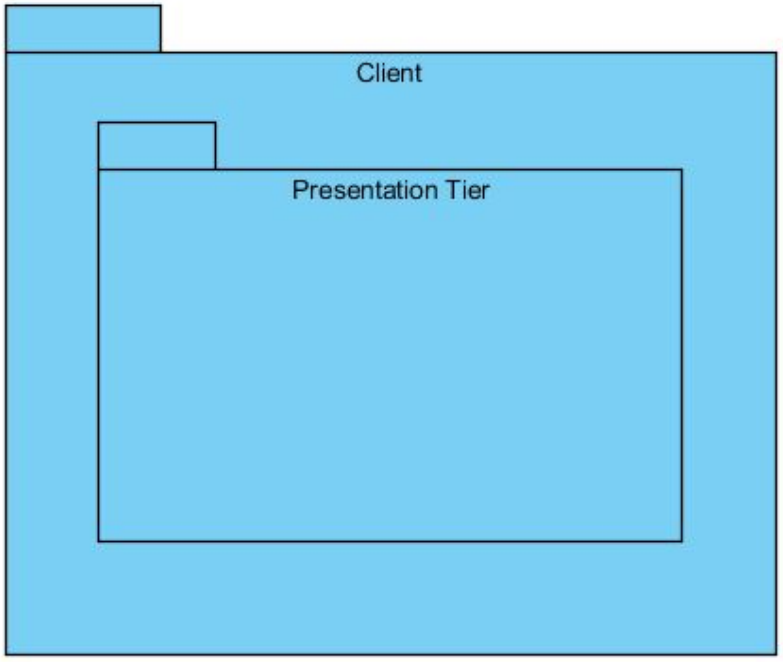
2.1.Subsystem Decomposition

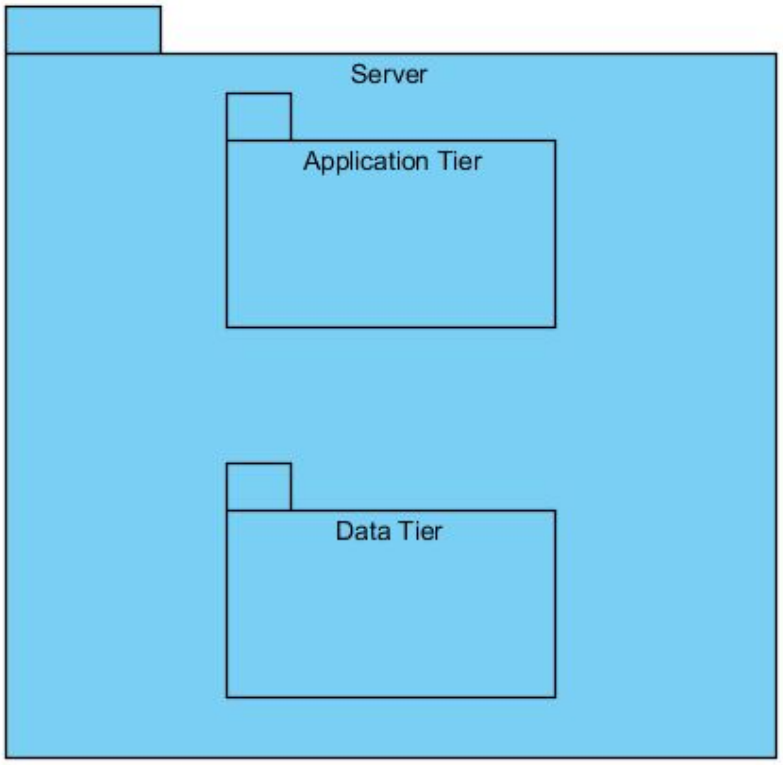
DAOS mainly follows a Client/Server architecture where we designate the server to handle all queries and calculations. The client side provides the interface to our user to communicate, make queries and changes to the server and display the response from the server. Client will be a single executable on windows which present all of our functionalities to the users. Many clients will connect to our one server which is an apache HTTP server working with a MySQL database solution. This way we aim to provide smooth transitions on our interface to our user provided by the client and all the processing load is on the server. Maintenance and performance increase of the server can be made centrally so scalability is ensured with this architecture model.

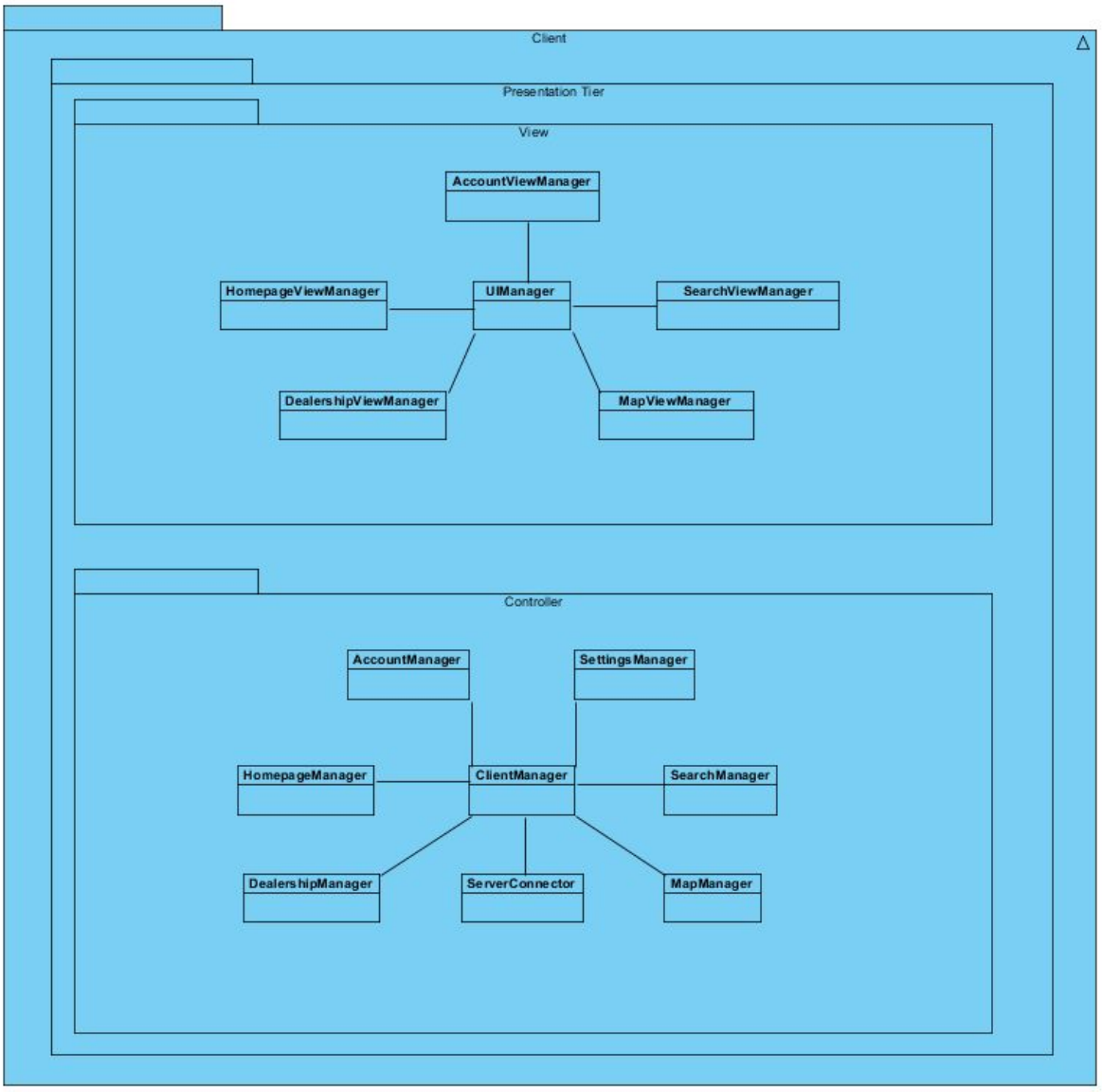
Client has only presentation tier which consists of a View subsystem and a Controller subsystem. View is mainly responsible for UI and display. The user consistently interacts with the view subsystem and this subsystem is responsible for transmitting the inputs to the controller and provide the user with a seamless to use interface. All subcomponents of the view subsystem is connected to a main component UIManager and every compartment interacts with another via UIManager. The controller subsystem is responsible for supervising the incoming commands from the view and transmit them via ServerConnector comportment. The main compartment in the controller is ClientManager and all the other ones interact with ClientManager then ClientManager preprocessing is complete the instructions to our server is made by ServerController which can work bidirectional which means both the user input and the result created by our server passes through this route.

Server has two tiers, Application tier and Data tier. Application tier is responsible for all the processing which will be performed by DAOS the inputs come into application tier with ClientConnectionManager (which itself communicates with ServerController in the Controller in the Client side). This is the component where the input or query comes then be processed

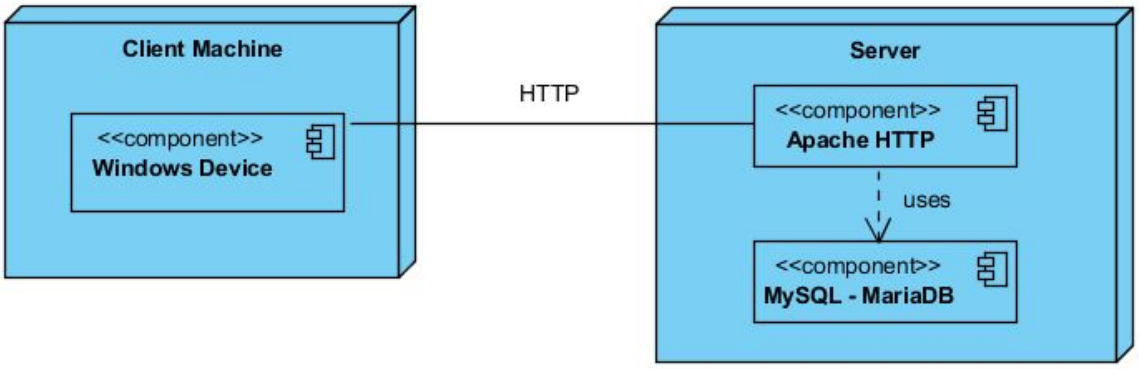
or the result of the query's final destination in the server side. Application tier also communicates to data tier which all of our information is stored in the database for apply the logic of the DAOS and if necessary make the updates on the database.






Client
Presentation Tier
View
AccountViewManager
HomepageViewManager
UIManager
SearchViewManager
DealershipViewManager
MapViewManager
Controller
AccountManager
SettingsManager
HomepageManager
ClientManager
SearchManager
DealershipManager
ServerConnector
MapManager

2.2.Hardware / Software Mapping



DAOS requires an internet connection in order to connect to the database. Client machine will interact with the server using HTTP requests. When a user opens the program on their computer it will be connected to the MySQL database which is in Apache server.

2.3.Persistent Data Management

Time is one of the most important things for companies. We need to provide information as soon as possible. Some of our objects such as person and dealership will not be changed if there is no need to change information about dealerships or the person using DAOS. Until the user change these informations, adds or removes dealership, the data remains. If the user searches dealerships, it needs to be shown quickly. We chose to store our data in MySQL - MariaDB, we believe it will satisfy our needs.

3.DAOS Algorithmic Design

3.1.Data Collection

The collected data is from the Adana province of turkey. We are holding data for the entire country of Turkey but the province selected for testing and fine tuning for our platform is performed on Adana only for the first version. The data consists of 410474 lines. Each line representing a square which is 30 meters to 30 meters in Adana. This squares are in a grid geometry but we omitted the squares where no residential area is possible (i.e. lakes, mountains etc.) since a shop cannot be opened in the middle of nowhere. This square

representation is selected to improve the precision of DAOS regarding the recommendations of the platform. We observed that recommending a neighborhood for a branch still covers a lot of area which needed to be narrowed down even more. All area features have many attributes about roads, buildings, population density, population, weight of the population compared to the region it resides on etc.

3.2.Feature Selection

There were nearly 50 features for each datum but the ones which will be useless are not fed into the learning algorithms such as registration no, id, name of the county, name of the neighborhood and name of the province(since recommendations are made within a user selected province) etc.

We selected 9 features that are most important when determining the type of the recommendation. The features are mean population density which is a number, clutter type represents the type of the settlement and the foot traffic within that square, population density inside the square, maximum people, resident population, types of the buildings (each will be a separate feature instead of a composite single feature. Total of 14 consecutive binary indicators), a binary indicator if the square includes a road or a street, population of the neighborhood the square is in, and the binary indicator for the already functioning shops. The last feature can have 15 different values and one square can have more than one shop. So there will be 14 different data sets to fine tune our predictions even more. The food dataset will use the first eight features as it is but the last feature will be set to one if there is an existing food shop in that area. And this is similarly implemented for the other types of shops for each county. Our user interface passes selections of province and county information to our recommendation subsystem and it produces all the possible venues in that county.

3.3.Learning and Recommending

3.3.1.Opening Recommendations:

The opening recommendations are made on the go with Random Forest classification algorithm. The user interface will pass the selected county area and the dealership category as parameters to the recommendation system. Then the recommendation system will use every other county but the selected one to train then it will predict the ones in the selected county dataset with all of them are labeled according to same shop category. Data is divided into directories which are mapped to the counties and

each directory has 14 files containing the squares inside that county. Every file has labels according the filename which is the category of the shop. This predictions then are passed back to our user interface.

3.3.2.Closing Recommendations

This recommendations will rely heavily on the client given sales data. This category also incorporates the machine learning algorithms developed by us but the influence weight is very small compared to the opening recommendations.

3.3.3.Tuned Hyper Parameters for Random Forest Classifier

All hyper parameters are set to their default values except the ones listed below. No additional fine tuning made to avoid overfitting. Mean accuracy with data split as 80% train and 20% test is between 98% - 98.5%.

n_jobs: is set to -1 to parallelize the jobs the count of the cores of the system. This compensates the storage access latency for our complete dataset which is already huge for a single city.

max_features: The number of features to consider when looking for the best split was set to auto at first then it was set to none which yielded a score slightly better and the performance tradeoff was not so impactful so no limit was imposed to max_features.

n_estimators: The number of trees in the forest was set as high as it can be while keeping the performance tradeoff as small as possible. Which turned out to be 30.

min_samples_leaf: The minimum number of samples required to be at a leaf node set to 25 with the help of a for-loop.

4.Impact of Engineering Solutions Developed in the Project

DAOS has many global, societal and economic impacts as all engineering solutions.

4.1.Global Impact

Many companies use their own developed performance analysis, assistance and forecasting programs in order to make better decisions, prevent loss of money and see the future clearer. To develop and to maintain that kind of program those companies need to hire qualified employees. They often need to create a whole new department responsible for it. For big companies, this is not a problem but for smaller companies, it becomes harder to

finance such a department and employees. In the market, there is not an application what DAOS presents.

DAOS provides user with its assistance feature by making suggestions for new dealerships' place and DAOS ensures that the user is informed by the advanced performance analysis and by the dynamic GUI about existing dealerships. Basically, DAOS is an application for the companies that are need of a number of dealerships to sell their product. This necessity makes DAOS a useful, money-winning, wanted and impactful application.

4.2.Societal Impact

DAOS is not only a useful application for companies but it is useful for society as well. Many people need to travel long distances to reach the service they are looking for, to buy from a specific brand or even to play the lottery and what is worse, many people don't know about the brand or lottery because of the absence of dealerships around. DAOS aims to suggest the best places for dealerships where more people can take advantage of the service. While provides company with more customers, customers will have more choices.

4.3.Economic Impact

DAOS can have a huge impact on economics since it can be used by any kind of business that requires dealerships. Using DAOS, companies will find the best places to reach their customers and they will gain a good amount of new customers by opening new dealerships on the most efficient places. Companies will profit from DAOS not only by gaining new customers but also not investing on wrong dealerships because every dealership is an investment for future that can make money or lose it. DAOS use a sophisticated algorithm that considers every location for a business by using GIS and makes decisions by using machine learning. That reduces the risk of an investment and increases the probability of a right choice. Employees of companies are of course the expert of their job. DAOS's well designed and dynamic GUI and performance section will help them to keep track of performance of all the dealerships they have. This will positively increase employees' performance.

5.Contemporary Issues Related with the Area of the Project

One of the main issues we were faced with in our project was the difficulty in finding data . DAOS depends on data about cities and their counties such as popuşation density, clutter type of the settlement, foot traffic resident population and types of buildings. We rely on all these data in our recommendation algorithm.

DAOS uses Google Maps to show recommended areas. Google Maps needs to be available when user uses DAOS system .Therefore, if Google API services fail, users cannot be provided with location image.

Security is also one of the most important issues. The user can login to his system with his email and password. Hence, the security is crucial for DAOS. We strengthen the storing way of user's information to prevent those kind of problem.

Internet connection is crucial for the functioning of the system. The application relies on internet connection since it is an web application. Hence, in case of a failure of internet connection, the application may be out of service temporarily.

6.Tools and Technologies

PyCharm is the python IDE [1].
Anaconda is the python distribution [2].
Global Mapper is the GIS software [3].
GitHub online Git version control system.
MySQL for the database system.
Google Maps API for showing recommended location.
NetBeans IDE for GUI implementation.
Apache Tomcat for servlet container
Apache Web Server for MariaDB database
JCreator for database interaction

7.Resources Used

7.1.Library Resources

Scikit is the machine learning library we used [4].

Random forest classifier is the machine learning algorithm [5].

Primefaces

MariaDB-SQL for Java

7.2.Web Resources

https://stackoverflow.com/
https://developers.google.com/maps/
http://www.hyperlinkcode.com/
https://www.w3schools.com/

8.Similar Products

There is not any current system for dealership assistance and performance review. Companies may use similar systems designed by their IT department but any of them is public.

9.Possible Future Development

There are some possible future developments for DAOS.

Firstly, we just focus on the Adana, but we want this application throughout the Turkey. Also when DAOS makes recommendations it does not uses the actual geometry and relations of the recommendation points between them. We are planning to make recommendations calculated ad hoc for every client with incorporating their sales data. In addition with more data we can also tune our hyperparameters even more. Client provided data will allow us to develop our closing recommendations even more.

10.User Manual

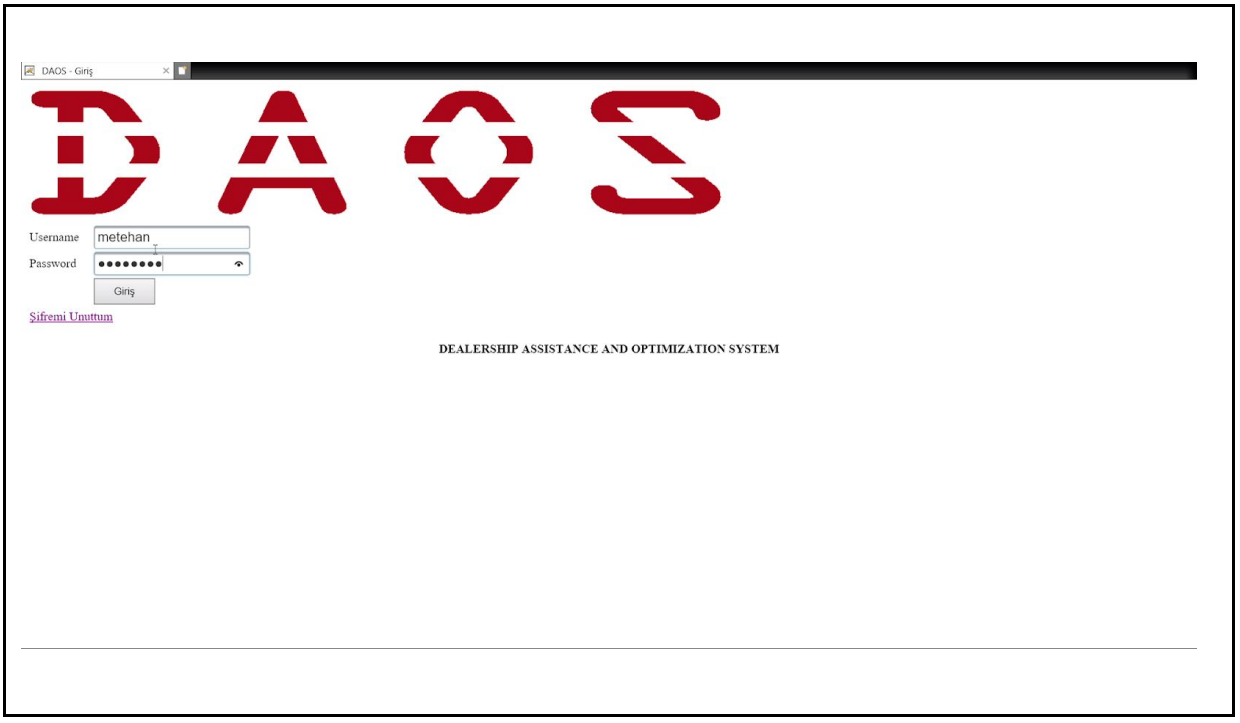


Fig 10.1 Login Screen

Login page is the first page to be seen. User enters username and password to login. If username and password are filled correctly, then the program will direct user to homepage. There is also a “Forgot my password” button in login page. If user clicks this button the program will direct user to “Forgot my password” page. If any login information is wrong, a small error message with try again will be presented to user.

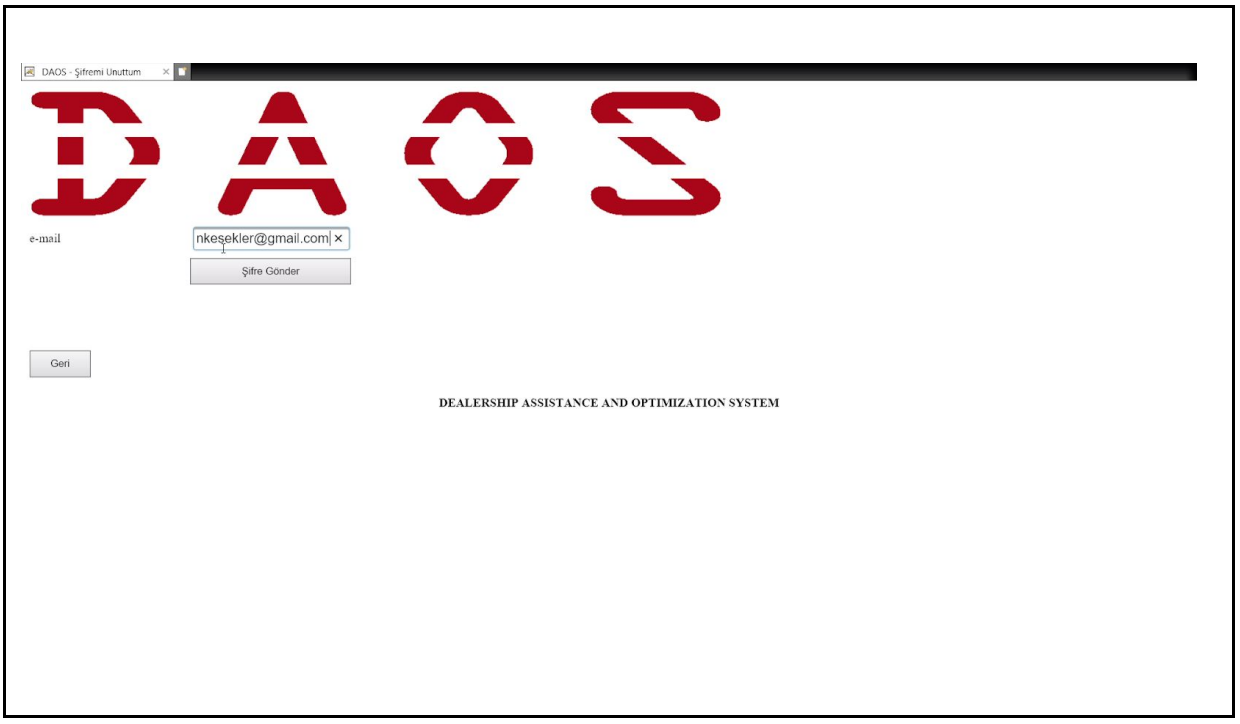


Fig 10.2 Forgot Password

When user clicks “Forgot my password” this screen appears. User can enter his email and get his new password. If user back button the program will direct user to login page.

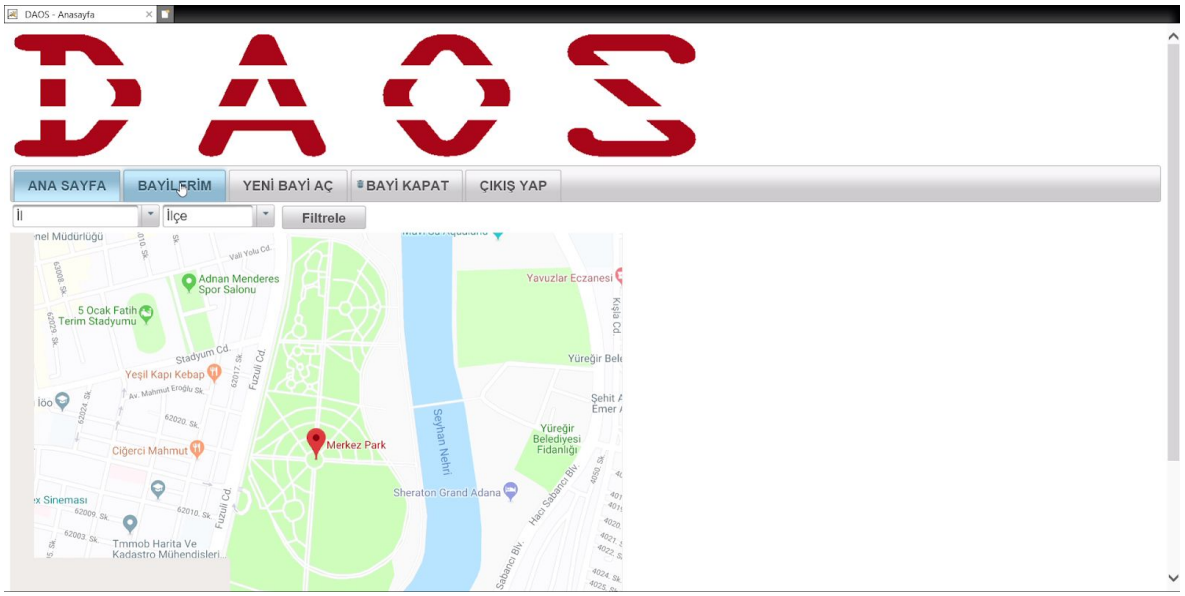


Fig 10.3 Homepage

Homepage of the program. When user logins the screen that takes place above will be shown.

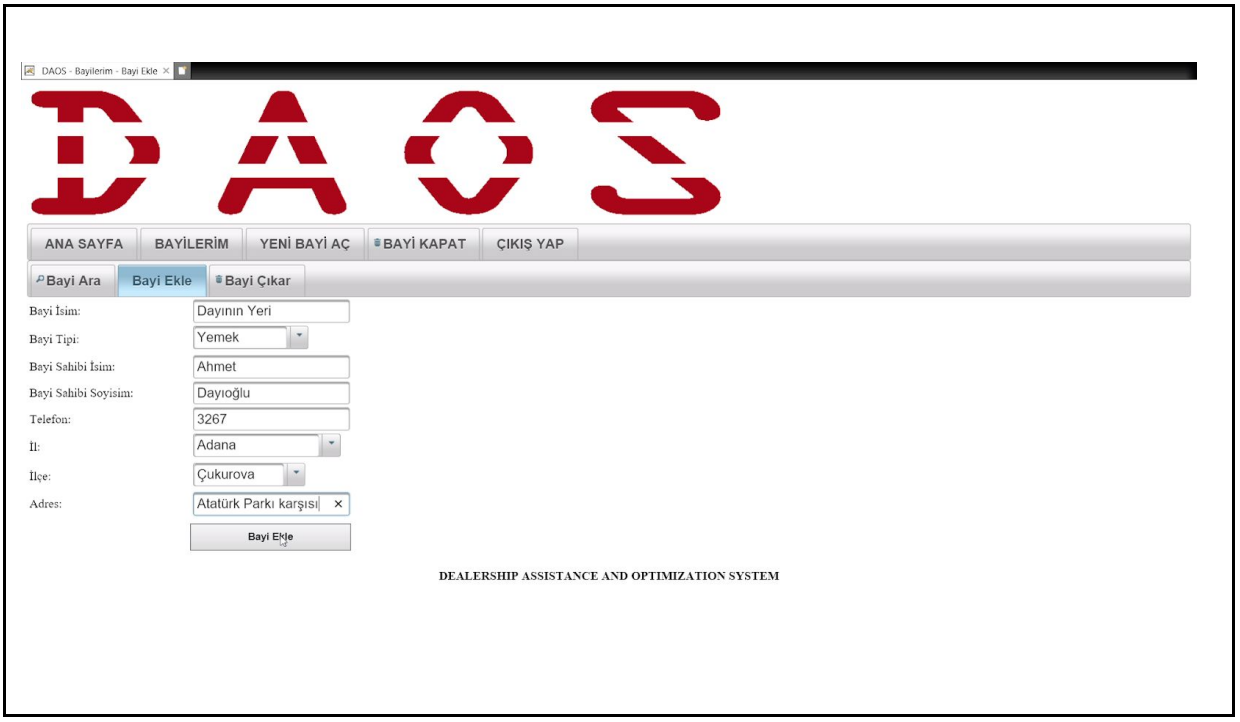


Fig 10.4 Add Dealership

When user clicks “Add Dealership” he will be able to enter the necessary information for the dealership and its owner to add a dealership.

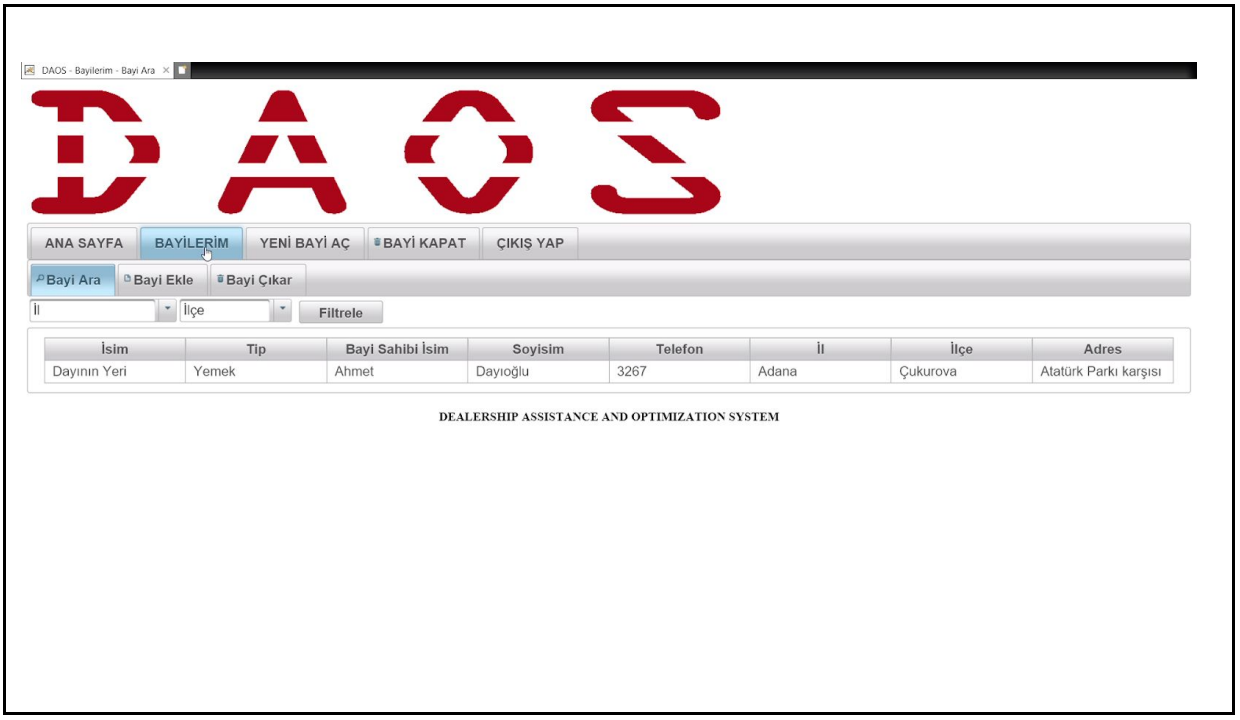


Fig 10.5 Search Dealership

When user clicks “Search Dealership” he will be able to filter city and county to see all dealerships he has on those areas.

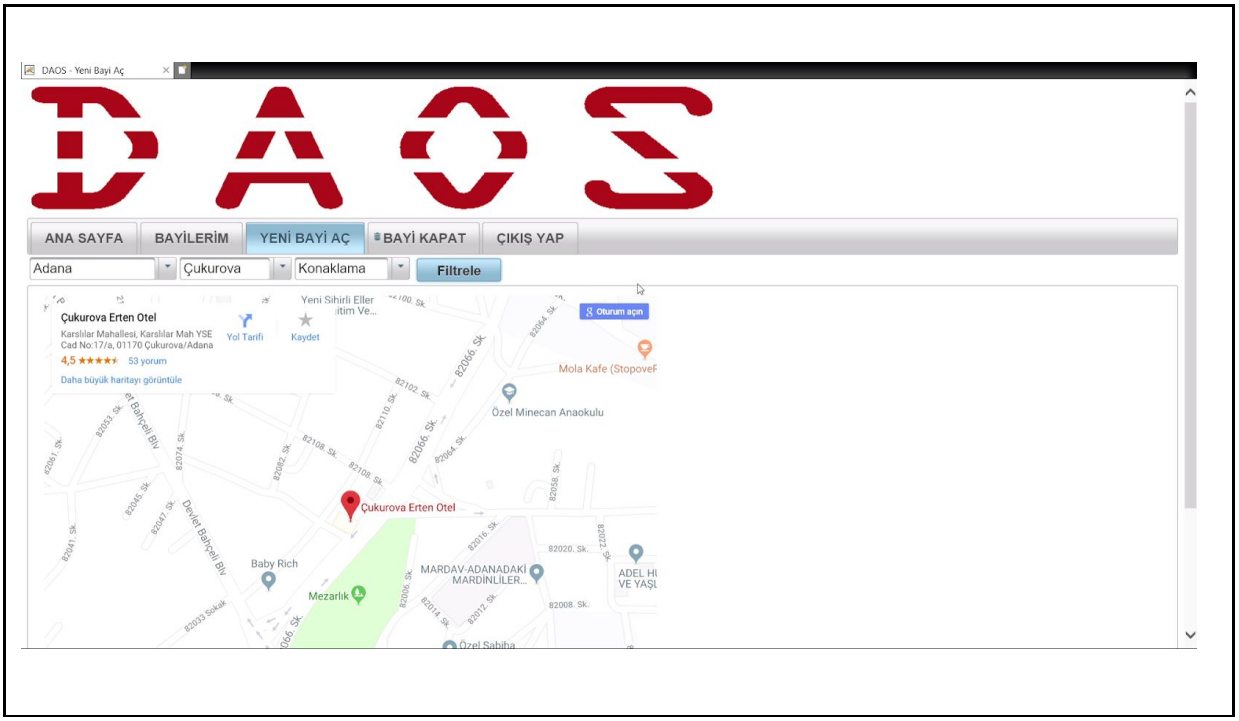


Fig 10.6 Open Dealership

When user clicks “Open Dealership” on the homepage he can filter the areas he wants to open a new dealership and the program suggests a location to be opened.

11.References

[1] “PyCharm: Python IDE for Professional Developers by JetBrains.” *JetBrains*, www.jetbrains.com/pycharm/.

[2] *Python :: Anaconda Cloud*, anaconda.org/anaconda/python.

[3] “Global Mapper Overview.” *Global Mapper - All-in-One GIS Software*, www.bluemarblegeo.com/products/global-mapper.php.

[4] “Sklearn.ensemble.RandomForestClassifier.” *3.2.4.3.1. Sklearn.ensemble.RandomForestClassifier - Scikit-Learn 0.19.1 Documentation*, scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[5] “Scikit-Learn.” *Scikit-Learn: Machine Learning in Python - Scikit-Learn 0.19.1 Documentation*, scikit-learn.org/stable/.