PROJECT 2

Submit your project in *WISEflow*. The submission deadline is Monday October 28th, at 12:00 hr (noon). The project can be done individually or in a group of at most 2 students. No cooperation between people who are not submitting this project as a group is allowed. It is possible to change groups throughout the semester and it is also possible to do some project(s) alone and other project(s) in a group. Provide **all your AMPL files** (model code, data, running commands, solution file, etc.) compressed in a single file (.zip). Include all files needed to run all parts of the project, even if from one to another task the changes are just marginal (we need all files to be able to run without modifying what you submitted). In addition, provide a written report as a **pdf file with your model formulations** and the answers to the questions required in each part. The formulation of your models can be typed in a text editor (e.g. Word, LaTeX), written by hand and scanned, or copied directly as text or screenshot from the AMPL code files when it applies (please just be careful the presentation must be clear enough for a reader). In the written report, it is fine that when there is just a marginal change from one task to another, in the latter you include just the modified part of the formulation (e.g., in task 2 you just defined a new variable or modified one constraint of the model you formulated in task 1, then it is fine that you included the full model formulation in task 1 and only the new variable definition and new constraint that you modified in task 2). When asked to solve a model, use the solver *gurobi.* Provide a short description (no more than a few words, e.g. "demand fulfillment") for every objective function and constraint in your formulations. Only as a reference (not as a requirement), the expected length of your report is: Part A one page, Part B five pages, Part C three pages. All model formulations in this project, either involving continuous and/or integer variables, must be **linear**.

Part A (20 %)

In the article "Volkswagen Group Logistics Applies Operations Research to Optimize Supplier Development" (available in the *Complementary readings* folder in *Canvas*), the authors describe the application of several models to improve the logistics performance of the supply network of a major automobile manufacturer. In the following tasks, we focus on two of the models formulated in the Appendix of the article. Each task requires you to formulate some modifications to these models to capture some new situations. Your formulations must be linear and may involve new definitions (e.g. of variables), new expressions (e.g. in the objective function and/or constraints), modification of some expressions in the original formulation, etc. These must be formulated in mathematical terms (not in AMPL code). The new situations are described below (each situation is independent from each other).

1. Tasks a) and b) below refer to the model formulated in the section "Capacitated Vehicle Routing Problem (Stage 2)" of the Appendix (page 158 of the article).

   a) Suppose that instead of minimizing distance, the objective is to minimize the overall CO2 emissions to serve all suppliers of the network. In practice, an accurate computation of emissions may be quite difficult, depending on a variety of factors such as the travelled distance and the freight volume carried by the trucks over each segment of their routes. We adopt here a simple approach, assuming that the company has estimated emissions based on past data and expert support. In this approach, the total emissions of a truck equals an amount $\bar{E}$ if the total demand in the tour of the truck is more than a threshold volume t. Otherwise, the total emissions of the truck is equal to E (where $\bar{E} > E$). The values of these parameters are given to you as data. Notice also the parameters defined in the article are known to you, including b_i, the freight volume demanded at supplier i. Which modification(s) would you introduce in the model to capture this situation?

b) Suppose that some of the roads between suppliers are quite congested, which often result in *stressful* journeys for the drivers of the trucks. For all supplier i (including the depot), let us define N_i as a subset of S, which is given tou you as data. If j belongs to N_i, it means that the road to drive from supplier i to supplier j is a *stressful* road. The company has K drivers (that is, the same as the number of tours computed by the model). Let us assume that $K > 5$. The company knows that only the five most experienced drivers are willing to drive tours with three or more of these stressful roads. Therefore, the solution to the model should make sure that no more than five trucks travel a tour including three or more *stressful* roads. Which modification(s) would you introduce in the model to capture this situation?

2. This task refers to the model formulated in the section "Knapsack Model (Stage 3)" of the Appendix (page 159 of the article). Note that Table A.1 reports some inaccuracies of the objective function, as it tends to overestimate the total benefit. This may relate to the network effects due to interdependencies between suppliers, especially when they are situated close to each other, as discussed on page 149 of the article. In order to correct this issue, let us define a set L containing tuples (j, k, h) such that suppliers j, k and h are relatively close to each other. For all tuples (j, k, h) in set L, let us assume that if measure i is allocated to suppliers j, k, and h, the objective function should decrease by an amount $u_{j,k,h}$. This amount and the set L are given to you as data. Which modification(s) would you introduce in the model to capture this situation?

Part B (40%)

One of the main international football events is the *UEFA European Football Championship* (or simply *Euro*). This is a quadriennial competition organized by UEFA, gathering 24 men's national teams from Europe. The latest edition was recently held in Germany, during the summer of 2024, featuring 51 matches accross 10 different cities. It has been reported that the total attendance to the stadiums was about 2.6 million people. While the event is undoubtedly exciting for football fans, it also involves a significant carbon footprint. In fact, not only fans, but also journalists, players, and administrative staff travel intensively across the host cities during the dates of the tournament. Although UEFA devised a series of sustainability measures at Euro 2024, one of the main authorities of the competition stated that "developing the schedule is still a very manual process... There is no computer automation involved"[1]. This motivates us to develop a mathematical programming approach to find a more sustainable schedule of matches for the group stage of Euro 2024.

In the group stage of Euro 2024, the 24 participating teams are divided into six groups of four teams each. Each team must play exactly one match against every other team within its group. Thus, in total there are 36 matches during the group stage. Each match must be played in one of the ten cities hosting the competition: Berlin (BER), Munich (MUN), Dortmund (DOR), Stuttgart (STU), Gelsenkirchen (GEL), Hamburg (HAM), Düsseldorf (DUS), Frankfurt (FRK), Cologne (COL), and Leipzig (LEI).

The group stage matches must be scheduled over an interval of 13 days, spanning from 14 June to 26 June, 2024. Figure 1 shows the actual schedule of games and the venues to which they were allocated in the solution used by UEFA.

In our modelling approach, we will attempt to retain many of the characteristics of the real solution. First, we should respect the exact date in which each of the 36 matches was played. For example, as Italy played against Albania on 15 June, our schedule should also feature this match on this date. Note that by retaining this characteristic, we take care that the teams will have exactly the same number of rest days between consecutive games as they had in the actual competition and that the number of matches per day is exactly as scheduled by UEFA (which might be important to satisfy agreements with sponsors and TV broadcasters).

We should also take into account important considerations about the venues. First, each of the 10 venues should host a number of matches equal to the number of matches it hosted during the group stage of the actual competition. For example, as shown in Figure 1, the venue in Berlin hosted three matches and the venue in Hamburg hosted four matches. Therefore, our solution should also feature three matches in Berlin and four matches in Hamburg. Also, due to maintenance on the pitch and the deployment of

[1] *Sustainability key to EURO 2024 schedule*, https://www.uefa.com/euro2024/news, June 12, 2024.

infrastructure, we should make sure that after a match takes place on a venue, there are at least two calendar days without a match at this venue. This is also a characteristic of the actual schedule, as can be seen in Figure 1. For example, after the match between Portugal and Czech Republic in Leipzig on 18 June, there are two calendar days (19 and 20 June) without a game in Leipzig, while after the game between Turkey and Georgia in Dortmund on 18 June, there are three calendar days (19, 20 and 21 June) without a game in Dortmund.

In what respect to time, it is important to remark that our modelling approach is concerned only about calendar dates, and not about the exact time at which the matches are scheduled.

To reduce travels, it would be positive to keep the teams of each group around just few venues. On the other hand, the host cities would like to see some renewal of visitors, while the fans following their teams would like to take the opportunity for sightseeing at more than one host city. Taking into account all of these perspectives, we will require that the six matches of each group must be played in at least four different venues. This feature is also satisfied in the actual schedule, as shown in Figure 1. For example, the matches of group A are hosted by four different venues, while the matches of group B are hosted by six different venues. In addition, we will require that each venue must not host more than two matches of the same goup. As shown in Figure 1, this is satisfied by all venues in the actual schedule, where for example Berlin hosts one match of Group B and two matches of Group D.

To secure that all venues see some action early in the group stage, we will require that by 18 June (inclusive) all venues should have hosted at least one match. Note this is satisfied in the actual schedule, with Leipzig being the last city to host its first match, precisely on 18 June. Likewise, to keep activity in every city towards the end of the group stage, we will require that the latest match hosted by every venue should not be before 24 June. Again, this is satisfied in the actual schedule, with Leipzig and Düsseldorf being the first cities where the group stage finishes, both hosting their latest match on 24 June.

We will also consider criteria on the *main-seeded* teams. Based on the overall European Qualifiers rankings at the time of drawing the groups, the main-seeded teams were: Germany, Spain, England, France, Belgium, and Portugal. We will require that the first game of the main-seeded team of each group must be played at the same venue as it was in the actual tournament. Thus, for example, the opening match between Germany and Scotland on 14 June must be in Munich, while the match between Spain and Croatia on 15 June must be in Berlin.

1. Formulate an integer linear programming model to find a match schedule for UEFA Euro 2024 satisfying all requirements described above, such that the total distance travelled by the teams is minimized. Note we consider here that the distance travelled by a given team consists of the distance between venues of its first and second matches plus the distance between venues of its second and third matches[2]. For this purpose, the file "dataB.dat" contains a matrix with data on the driving distance between venues, expressed in kilometres. The data has been collected from the *Openroute service*[3]. A cell in the matrix contains the distance to travel from the venue in a row to the venue in a column. Note the matrix of distances is not symmetric. For example, the distance from Berlin's venue to Munich's venue is 571 km, while the distance from Munich's venue to Berlin's venue is 570 km.

   Implement and solve the model in AMPL, using the solver ***gurobi***. Where does Germany play its matches in your solution? How much is the total distance travelled by the teams in your solution? How does it compare to the 14773 km travelled in the actual UEFA's schedule (see last row in Table 2)? Assuming that 2500 cars are driven by the fans of each team and that the CO2 emissions per kilomter are on average 123.2 grammes, how many kilogrammes of CO2 emissions are saved by your solution with respect to UEFA's solution?[4]

[2]In practice, teams normally stay at base camps and perform return trips between their camps and the venues. However, this decision is mainly up to each team, while the match schedule decision is made centrally by UEFA earlier than the team base camp selection. Furthermore, national fans of each team would normally travel from venue to venue, which is the relevant distance we consider in this problem.

[3]https://openrouteservice.org/

[4]Data from Germany's Federal Motor Transport Authority (KBA) reports that the average CO2 emissions for passenger cars is about 123.2 grammes per kilometre (https://www.tagesschau.de/). Our estimation uses this number and it assumes that about 10,000 fans per team travel by car and that they do it in groups of four passenger per car. Of course, in practice the number of fans may vary across teams and many of them might have used other transport means and different routes, but we are just making a rough estimation here to put in perspective the environmental impact of our optimization approach.

In your written report, outline the match schedule that you found, including the pair of teams playing each match, and the corresponding dates and venues. This can be displayed graphically (e.g. drawing a figure similar to Figure 1), as a table, commented in words, or by any other means that could easily be understood by an external person. (Imagine you are communicating the solution to UEFA authorities, who are not necessarily familiar with mathematical programming, so they want to clearly understand the final solution instead of seeing a raw display with the optimal value of your variables.)

In tasks 2 and 3 below, we will not require the condition on the main-seeded teams. However, taking into consideration that Germany was the host team, we will require that the three group stage matches of Germany should be played in the same venue as in the actual schedule. Thus, Germany vs Scotland on 14 June should be played in Munich, Germany vs Hungary on 19 June should be played in Stuttgart, and Germany vs Switzerland on 23 June should be played in Frankfurt. In fact, we might understand that the venues of the matches featuring the host team were predefined early in the match scheduling process. Furthermore, the host cities likely played an important role in the allocation of matches to their venues. In particular, the dates of the matches hosted by every city are important for local authorities to plan security and local transport measures. Therefore, in all tasks that follow we will require that every venue must host one match in every date where the actual schedule indicates so. Looking at Figure 1, for example, Gelsenkirchen should host a game on 16 June, on 20 June and on 26 June. All other requirements remain the same as we described previously (except for the requirement about main-seeded teams).

2. Modify your model (keeping linearity) to find a match schedule for UEFA Euro 2024 in this new situation, such that the total distance travelled by the teams is minimized. Implement and solve the model in AMPL, using the solver ***gurobi***.

   a) How much is the total distance travelled by the teams in your solution? How does it compare to the 14773 km travelled in the actual UEFA's schedule? Using the estimations as we did in task 1, how many kilogrammes of CO2 emissions are saved by your solution with respect to UEFA's solution? In your written report, outline the match schedule that you found, using the same format as you did in task 1. Where do the main-seeded teams play their first games?

   b) Among all 24 teams in your solution, which team is the one that travels the longest distance? Which team is the one that travels the shortest distance? How many kilometers each of these two teams travel? How much is the difference between their travelled distances? We may interpret this difference as a measure of *fairness*. The lower the difference, the fairer the schedule. Note in the actual solution by UEFA, the team that travels the longest distance was Ukraine, with 1075 km, while the team that travel the shortest distance was Switzerland, with only 196 km. Thus, the maximum difference of the travelled distances among all pairs of teams in the group stage of Euro 2024 was 879 km. In BAN402, maybe we can do better...

3. The paragraph above motivates us to define a new objective, which is to minimize the maximum difference of the travelled distances among all pairs of teams in the group stage of Euro 2024.

   a) Modify your model (keeping linearity) to find a match schedule for this new situation. Implement and solve the model in AMPL, using the solver ***gurobi***. In your solution, which team is the one that travels the longest distance? Which team is the one that travel the shortest distance? How many kilometers each of these two teams travel? How much is the difference between their travelled distances?

   b) Can you find a match schedule that fulfils the optimal travelled distance obtained in task 2a and also the optimal difference obtained in task 3a? If yes, outline such a schedule and indicate for which pair of teams is the longest difference observed.
   If not, what is the optimal difference that you obtain while fulfilling the optimal distance obtained in task 2a? If we allow for a total travelled distance of at most 5% larger than the optimal distance found in task 2a, what is the optimal difference? What if we replace 5% by 10%?

Note: Remember to include the model formulations in your report.
When solving mathematical programming models, it is interesting to observe how the algorithms of the solvers approach the optimal solution during the optimization process and some statistics on the dimension of the problem and the solution time. For this purpose, when using *gurobi*, you can add the following lines in your .run file anywhere before the statement *solve*:
option solver gurobi;
option gurobi_options 'outlev=1';
option show_stats 1;

Part C (40%)

The company BanPetrolytics buys several types of crude oils daily at its two refineries (R1 and R2). The cost of purchasing one unit of crude oil i on day t is $C^{CRU}_{i,t}$. The crude oil is converted to components through a series of processes in the refineries. A crude oil which goes through these processes provides different components. The proportions in which these components are obtained vary from one to another refinery, because of the different technology and conditions under which the conversion occurs. We will refer by $a_{r,i,b}$ to the amount of component b obtained from processing one unit of crude oil i at the refinery r. The maximum processing capacity of crude oil per day at the refineries is 850 units at R1 and 810 units at R2.

The cost of processing one unit of crude oil i at the refinery r is $C^{DIS}_{r,i}$. The crude oil is ready for being processed within the same day of purchase or, alternatively, it can be stored at the corresponding refinery. The cost of storing one unit of any type of crude oil is C^{INVI} per day. However, it is not possible to store components at the refineries. Therefore, all the components obtained from processing the crude oils are sent to another facility of the company that we will refer to as the hub. Assume that the components sent from the refineries on day t are received at the hub on day $t+1$, due to some lead time of transportation and handling. The cost of transporting one unit of any component from a refinery to the hub is C^{TRA1}.

Once the components arrive at the hub, they are ready to be mixed for generating final products. The mix occurs according to predefined recipes. The recipe for producing one unit of product p needs $Q_{b,p}$ units of component b. The cost of producing one unit of product p is C^{PRO}_p. It is possible to store components in the hub at a daily cost of C^{INVB} per unit. In contrast, the final products cannot be stored here and, instead, are sent to depots. The cost of transporting one unit of any product from the hub to depot d is C^{TRA2}_d. Assume that what is produced during one day arrives at the depots in the beginning of the following day.

Once the products arrive at the depots, they are ready to be shipped to the markets. The cost of shipping one unit of any product from depot d to market k is $C^{TRA3}_{d,k}$. Alternatively, the products may be stored at the depots. The cost of storing one unit of any type of product at depot d is C^{INVP}_d per day.

There is a maximum demand limit for product p in market k on day t, which we will denote by $\delta_{p,k,t}$. The price of one unit of product p in all markets is S_p. Assume that it is possible to fulfil demand of a same market partly from different depots. Also, assume that what is shipped from the depots one day arrives to the markets the following day, except for two markets in the extreme south and two markets in the extreme north which take an extra day. These markets are labeled as *ES1*, *ES2*, *EN1*, and *EN2*. The shipments to these extreme markets have to pay a fixed cost. If a positive quantity of products (in total) is sent from depot d to extreme market k on a given day, the fixed cost incurred is equal to $CExtreme_d$. Note there might be shipments from more than one depot and to more than one extreme market, and the fixed cost must be paid for each used link. Due to a commitment with local authorities, you have to assure that there is at least one day over the planning horizon with shipment to each extreme market. Also, due to a policy within the company, there cannot be shipments to the extreme south and to the extreme north on the same day.

Figure 2 illustrates the different stages of the supply chain of BanPetrolytics. The file "dataC.dat" contains data for this problem. You may modify it according to your own formulation. Note the set T contains 11 days, including from day 0 to day 10. The day 0 is only relevant for initial conditions (you may eventually delete it from the set T in the data file if your code does not need it). Note demand on day 10 is satisfied by shipping on day 9 (or day 8 in case of the extreme markets). It is not possible to make a profit from shipping products that would arrive at destination after day 10.

1. Formulate a mixed integer linear programming model for the company, including decisions on procurement, production, transportation, storage and sales (use only continuous and binary variables). The objective is to maximize the total profit over the planning horizon. Assign value zero to any initial inventory or initial flow variable that you may require in your formulation, except for the initial inventory of product p at depot d which is equal to the data given in the columns under the header $Izero_{p,d}$ in Table 1. The company would like to have some final inventory of components and products (that is, the inventory at the end of day 10), in order to anticipate future demand. The final inventory of products for each depot is given in Table 1 under the header $Ifinal_{p,d}$. The final inventory of components *distilA* and *distilB* must be at least 100 units each, while for components *ISO* and *POL* it must be at least 400 units each. Note all the inventory costs are incurred per any unit stored at the end of each day (including day 10).

	$Izero_{p,d}$		$Ifinal_{p,d}$	
	D1	D2	D1	D2
premium	210	210	125	100
regular	400	470	200	250
distilF	54	100	50	30
super	200	135	30	60

Table 1: Initial and final inventory quantities.

Remember to include the model formulation in your report. Implement your model in AMPL and solve it using the solver *gurobi*.
a) What is the optimal profit? When do the shipments to the extreme markets start and from which depot(s)?
b) How much unsatisfied demand do you observe for each of the products (in total adding up all time periods and markets)
c) How the constraints on the minimum final inventory of components are satisfied in the optimal solution? Is there any difference in the corresponding slack variables? Why?

2. The prices of the crude oils in the data file "dataC.dat" remain constant over the days ($77 and $75 for *CrA* and *CrB*, respectively). Suppose an increase of the prices is forecasted to occur on day 6. According to this new forecast, the price per unit of *CrA* on day 6 and later will be $82, while the price per unit of *CrB* in these days will be $80. Solve the model for this new scenario. What is the optimal profit? How much are the inventories of crude oils stored at the end of each day? How do these compare to what you obtained for the original scenario?

Figure 1: Stadiums, groups, and actual match schedule of UEFA Euro 2024 (reproduced from https://editorial.uefa.com/).

Table 2: Groups, teams, teams' abbreviations, distance travelled, and an estimation of CO2 emissions, as for the actual UEFA Euro 2024 match schedule (a " * " next to the team name means this team was the main-seeded team of its group).

Group	Team	Abbreviation	Distance travelled (km)	CO2 emissions aprox. (kg)
A	Germany *	GER	422	129,976
	Scotland	SCO	991	305,228
	Hungary	HUN	378	116,424
	Switzerland	SUI	196	60,368
B	Spain *	ESP	569	175,252
	Croatia	CRO	689	212,212
	Italy	ITA	485	149,380
	Albania	ALB	754	232,232
C	Slovenia	SVN	835	257,180
	Denmark	DEN	622	191,576
	Serbia	SRB	669	206,052
	England *	ENG	471	145,068
D	Poland	POL	791	243,628
	Netherlands	NED	578	178,024
	Austria	AUT	565	174,020
	France *	FRA	908	279,664
E	Belgium *	BEL	576	177,408
	Slovakia	SVK	477	146,916
	Romania	ROU	809	249,172
	Ukraine	UKR	1075	331,100
F	Turkey	TUR	353	108,724
	Georgia	GEO	700	215,600
	Portugal *	POR	461	141,988
	Czech Republic	CZE	399	122,892
		Total	14,773	4,550,084

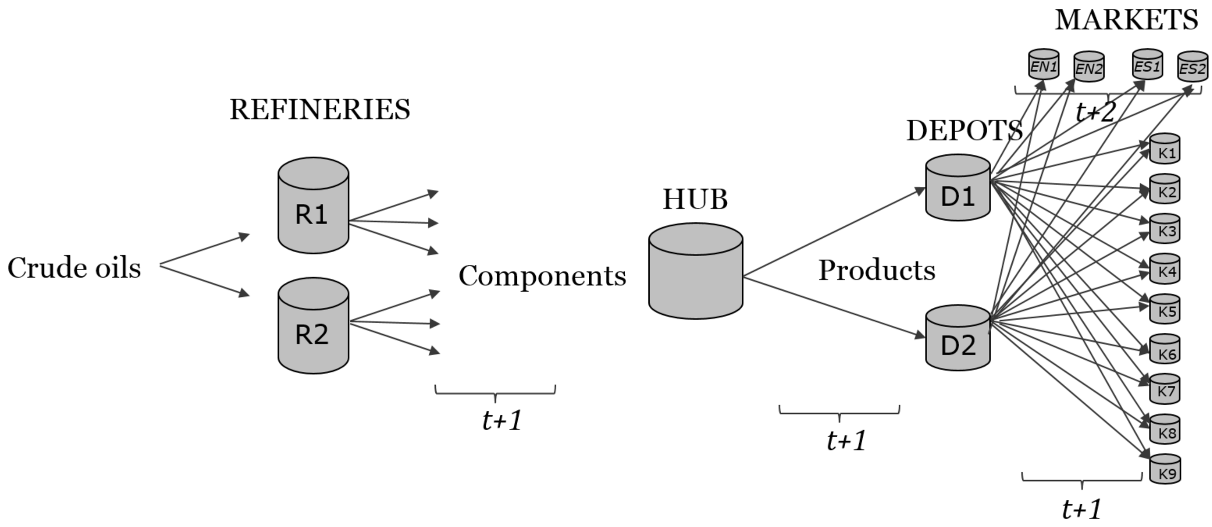

Figure 2: Illustration of the supply chain of the company.