Project Name	Sales Lead Qualifier
Online team meeting	https://fau.zoom-x.de/j/61899002491?pwd=WGFIL2tSS2lVWFhqZG9PWmMzVFg4dz09
Production system (if any)	(none yet)
Test system (if any)	(none yet)
GitHub repository	https://github.com/amosproj/amos2023ws06-sales-lead-qualifier
GitHub feature board	https://github.com/orgs/amosproj/projects/30/views/2
GitHub impediments backlog	https://github.com/orgs/amosproj/projects/31/views/1
Team T-shirt (white female)	https://www.shirtinator.de/s/PRS_LRm2Qx6vXMtaeB835A
Team T-shirt (white male)	https://www.shirtinator.de/s/-mNT0jHvSWGYH5w2QbGisQ
Team T-shirt (black female)	https://www.shirtinator.de/s/EIuAi2y0QS-VMak3P_gdjg
Team T-shirt (black male)	https://www.shirtinator.de/s/nVtVQvA1TjyCIkyXQhqNsw
Team T-shirt (all in shopping cart)	https://www.shirtinator.de/loadBasket/Wgfl6csWaoA
AMOS Happy	https://happy-amos.appspot.com/Project?project=4809731176660992&course=6210557241720832
Slack (team channel)	https://app.slack.com/client/T02J8GLTXDH/C061LNFLW13
Availabilites (Poll)	https://terminklick.stuve.fau.de/poll/AMOSP6/vote/
Extended Project Description	https://docs.google.com/document/d/1HpflwUUQqyQtFd1dMRjyoHYrDoZcacl_/edit

Goals	A deliverable product that satisfies the customers needs, Quality and Testing, User Satisfaction
	Apply best practice, deliver high quality code & documentation (by standards of the industry partner)
Meeting norms	Be punctual
	If attendance is not possible write a signal message to the team at least 1 day before.
	Find a balance for meeting/working time
Working norms	Be productive and do not have major crunch times.
	Automated testing, automated linting (to achieve uniform code style)
	Work with branches (bugs, features, ..)
	Work with development, staging and production stages
	Use technologies as defined by the industry partner (Programming language, AI framework, etc)
	Always review code before merging (by pull request)
Coordination norms	Work items are assigned to specific team members, the member feels responsible for its completion and to notify the team about issues
	Team members may form small groups / teams to work on components or tasks together
Communication norms	Communicate in case a meeting does not work out as early as possible, at least the day before
	Main channel is Slack, SD should communicate privately in case of specific issues
	Communicate technical issues in a corresponding slack channel of our team.
	Criticism should always be constructive, we do not tolerate bullying
Consideration norms	Discuss Issues with the team
	Transparency, providing clear information to customer
	Valuing customer's feedback
	Each team member's opinion is valued equally
Cont. improvement norms	Implement feedback of the team from previous weeks
Rewards	Everyone gets a drink of their choice (people in the same city can meet up) and post a picture in our group chat
Sanctions	A reason for being late on a meeting needs to be given.
Signatures	
Scrum Master	Nico Hambauer
Product owner	Simon Zimmermann
Product owner	Tetiana Kraft
Software developer	Felix Zailskas
Software developer	Ahmed Sheta
Software developer	Lucca Baumgärtner
Software developer	Resit Berkay Bozkurt
Software developer	Ruchita Nathani
Software developer	Fabian-Paul Utech
Software developer	Sophie Heasman

Last Name	First Name	GitHub User Name	Email Address	Private/ Google mail
Baumgärtner	Lucca	luccalb	lucca.baumgaertner@fau.de	kontakt@luccabaumgaertner.de
Bozkurt	Resit Berkay	rbbozkurt	resit.b.bozkurt@campus.tu-berlin.de	resitberkaybozkurt@gmail.com
Hambauer	Nico	nicohambauer	nico.hambauer@fau.de	
Heasman	Sophie	soapyheas	sophie.heasman@campus.tu-berlin.de	
Nathani	Ruchita	ruchita-nathani	ruchita.nathani@fau.de	ruchita.fau@gmail.com
Sheta	Ahmed	ultiwinter	ahmed.sheta@fau.de	ahmedhesham73@gmail.com
Utech	Fabian-Paul	ur-tech	f.utech@campus.tu-berlin.de	f.utech@gmx.net
Zailskas	Felix	felix-zailskas	felix.zailskas@campus.tu-berlin.de	felixzailskas@gmail.com
Zimmermann	Simon	Tims777	tim.simon.zimmermann@fau.de	

Sprint-#	Meeting Day	Product Owners	Software Developers	Release Manager	Scrum Master	Comment	Prior Assignment	Deliverable / Irr. Work	Amount RM
0	2023-10-18	Simon, Tetiana	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	N/A	Nico				
1	2023-10-25	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Lucca (x)	Nico			X	1
2	2023-11-08	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Berkay (x)	Nico			X	1
3	2023-11-15	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ruchita	Nico				1
4	2023-11-22	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ahmed	Nico				1
5	2023-11-29	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Fabian-Paul (x)	Nico			X	1
6	2023-12-06	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Felix (x)	Nico	Mid-term due		X	1
7	2023-12-13	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Sophie	Nico				1
8	2023-12-20	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Lucca (x)	Nico				2
9	2024-01-10	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Berkay (x)	Nico				2
10	2024-01-17	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Felix (x)	Nico		First Ruchita, then Lucca.		2
11	2024-01-24	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Fabian-Paul (x)	Nico		Ahmed, then Lucca		2
12	2024-01-31	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Sophie (x)	Nico		prior Fabian was assigned	X	2
13	2024-02-07	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ahmed (x)	Nico	Demo day!	Prior Felix was assigned	X	2
14	2024-02-14	Simon	Lucca, Resit Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ruchita (x)	Nico	Retrospective.	Prior Lucca was assigned	X	2
				x indicates the number of accomplished irregular works					

Product Vision	Project Mission
This product will give our industry partner a tool at hand, that can effectively increase conversion of their leads to customers, primarily by providing the sales team with valuable information. The modular architecture makes our product future-proof, by making it easy to add further data sources, employ improved prediction models or to adjust the output format if desired.	The mission of this project is to enrich historical data about customers and recent data about leads (with information from external sources) and to leverage the enriched data in machine learning, so that the estimated life-time value of leads can be predicted.

Term	Definition
Base Data Collector (BDC)	One of our main two software components which fullfills the task of collecting data about a lead from various online sources.
Expected Value Predictor (EVP)	The other main software component which takes the enhanced data about a lead and estimates the expected value of that lead.
Software Component	Major part of our product, which is mostly independent from other components (i.e. can be run separately).
Lead	Potential customer of SumUp which has already declared their interest in purchasing products or services (through the online form).
Controller	Software component that orchestrates BDC and EVP, leading them to an efficient collaboration.
Internal Data Source	Our industry partner providing us with their data about leads. This can be both previously collected or "live data" from new leads.
Provided Data	Data from the internal data source that serves as primary input for BDC. This data contains few features and is possibly irregular.
(Data) Feature	A common property among multiple data records (e.g. first / last name, phone number).
External Data Source	Third party providing us with data about a lead. In theory this can be the lead themselves (e.g. via information on their website).
Collected Data	All data about a lead, which has been collected from an external data source.
Enhanced Data	Combination of provided and collected data, possibly enriched by some derived features.
Output Data	Data which is produced by the EVP. It contains only features that are relevant for decision making from a sales point of view.
Expected Value	The revenue to be expected from a lead (expected value = life-time value of lead x probability of lead becoming a customer).
Lead Quality Index	An index in the range from 0 to 1, which is based on the leads expected value and can be used to rank or classify leads.
Data Label	A data label is the value of a data entry that the AI model should predict. In the training set they are set and known as ground truth label. In new data the model's goal is to predict the label. In our case the label is equivalent to the lead's value to SumUp.

Sprint	Sprint goal
1	Get Communication and Team Work going and establish a good working mode during meetings and with industry partner
2	Develop prototypes of BDC and EVP and possibly of the Controller, evaluate some first external data sources, do research on AI and AWS.
3	Establish further data sources and identify the challenges of collecting data. Put together all components and establish an automated build toolchain.
4	Improve current BDC data sources in tems of quality and quantity. Generate reports.
5	Continue work on BDC: Lay more focus on data-robustness and relevance for prediction. Use GPT to preprocess natural language data. Improve reports.
6	Prepare product for the upcoming sprint release: Stabilize, refactor and clean up code base
7	Catch up on documentation. Prepare to train useful AI model. BDC: Try collecting data for leads that we cant't find aynthing about yet.
8	Implement some more collection methods for derived data.
9	Preprocess historical data.
10	Use preprocessed data to train AI.
11	
12	
13	
14	
15	

Sprint	Goal	Feature Name	Est. Size	Est. Remaining	Real Size	Real Remaining
Release						
Total			120	120		
Sprints						
1	**Communication**		0	120	0	120
2	**Prototypes**		34	120	34	120
3	**Data Sources**		15	86	19	86
4	**Data Reports**		32	71	28	67
5	**Deeper Analysis**		19	39	15	39
6	**Stabilization**		17	20	17	24
	Sum			3		7
Features						
2	**Prototypes**	Create software bill of materials	2		2	
		Setup CI / CD pipelines	3		3	
		Implement EVP skeleton	5		5	
		Implement BDC skeleton	3		3	
		Data Field Definition	2		2	
		Implement Controller skeleton	8		8	
		Research different AI types	3		3	
		Research on external data sources	3		3	
		Setup experimental ML instance on AWS	5		5	
3	**Data Sources**	Automate build process	3		3	
		Use the Google Maps SDK instead of requests package	2		2	
		Find a safe way to use secrets in our project	2		1	
		Aggregate actual data	8		13	
4	**Data Reports**	X (Twitter) as Data Source	5		2	
		Create Reports	2		3	
		Fix Pipeline Demo input_location and output_location paths	1		1	
		Lookup of commercial e-mail adresses	3		3	
		Lookup of commercial phone numbers (in Germany)	3		2	
		AI model selection	3		3	

Sprint	Goal	Feature Name	Est. Size	Est. Remaining	Real Size	Real Remaining
		GPT as Data Source	5		3	
		Facebook & Instagram as Data Source	5		8	
		Database setup	5		3	
5	**Deeper Analysis**	Implementation of a Logger module	2		2	
		Retrieve business type	2		2	
		Download Reviews	5		3	
		AWS S3 Persistence Setup	3		3	
		Sentiment analysis	2		2	
		Fix redundant API Calls	5		3	
6	**Stabilization**	RegionalAtlas API as Data Source	5		5	
		Mock dataset	1		1	
		Fix bug in Google pipeline step	1		1	
		Create a default pipeline	1		1	
		Database Abstraction Layer	3		3	
		Implement review based quality metrics	3		3	
		Upload Reviews to S3	1		1	
		Prepare presentation	2		2	

Sprint	Goal	Feature Name	Est. Size	Est. Remaining	Real Size	Real Remaining
Release						
Total			90	90		
Sprints						
7	**Documentation**		12	90	7	90
8	**More Data**		8	78	10	83
9	**Data Preprocessing**		16	70	0	73
10	**Train AI Model**		0	54	0	73
11	**Improve Model**		0	54	0	73
12	**Unit Tests**		0	54	0	73
13	**Stabilization**		0	54	0	73
	Sum			54		73
Features						
7	**Documentation**	Document research results	3		3	
		Find an alternative for language_tool_python	3		2	
		Create location based qualtiy indicator	3		1	
		Document pipeline steps	3		1	
8	**More Data**	Find data for leads which we currently do not have data about	3		3	
		Fix for .env files	1		1	
		Let our pipeline run on historical data	1		3	
		Cache GPT results on S3	3		3	
9	**Data Preprocessin**	Simplify Reports	2			
		Possibility to amend to data	5			
		Use default pipeline during sprint release	1			
		Inspect lead "neighbors"	3			
		Preprocessing of collected lead data	5			
		Bug: Incorporating Regional Atlas doesn't let the pipeline run till the end				
10	**Train AI Model**	Train ML model prototype with historical data				

Sprint	Goal	Feature Name	Est. Size	Est. Remaining	Real Size	Real Remaining
		Business type analysis				
		Configure pipeline using a config file				
		Systematic approach to find more lead data (once again)				
11	**Improve Model**					
12	**Testing**					
13	**Stabilization**					

#	Feature Definition of Done	Sprint Release Definition of Done	Project Release Definition of Done
1	PullRequest has been reviewed, approved and merged	A release candidate has been created based on dev	*Same as sprint release but additionally:*
2	Done-Criteria in backlog item have been met where feasible	The release candidate and important features can be run	Relevant features have been documented
3	If criteria are not met, the team agrees that those should be dropped due to the specific circumstances	There is no major regression compared to the previous release (in the future, this should be tested automatically)	Entire code base has been revisited to identify overall product health, no major concerns have been brought up

Type	Link / reference
Markdown Files on GitHub	https://github.com/amosproj/amos2023ws06-sales-lead-qualifier/tree/main/Documentation
GitHub Wiki	https://github.com/amosproj/amos2023ws06-sales-lead-qualifier/wiki

#	Context	Name	Version	License	Comment	
1	pkg:pypi/requests@2.31.0	Requests	2.31.0	Apache Software License		
2	pkg:pypi/shapely@2.0.2	Shapely	2.0.2	BSD License		
3	pkg:pypi/beautifulsoup4@4.12.2	beautifulsoup4	4.12.2	MIT License		
4	pkg:pypi/email-validator@2.1.0.post1	email_validator	2.1.0.post1	CC0 1.0 Universal (CC0 1.0) Public Domain Dedication		
5	pkg:pypi/facebook-sdk@3.1.0	facebook_sdk	3.1.0	Apache Software License		
	pkg:pypi/geopandas@0.14.1	geopandas	0.14.1	BSD License		
6	pkg:pypi/googlemaps@4.10.0	googlemaps	4.10.0	Apache Software License		
7	pkg:pypi/numpy@1.26.2	numpy	1.26.2	BSD License		
8	pkg:pypi/openai@1.3.5	openai	1.3.5	-		
9	pkg:pypi/osmnx@1.7.1	osmnx	1.7.1	MIT License		
10	pkg:pypi/pandas@2.0.3	pandas	2.0.3	BSD License		
11	pkg:pypi/phonenumbers@8.13.25	phonenumbers	8.13.25	Apache Software License		
12	pkg:pypi/pydantic@2.5.2	pydantic	2.5.2	MIT License		
13	pkg:pypi/pymongo@4.6.0	pymongo	4.6.0	Apache Software License		
14	pkg:pypi/pytest@7.4.0	pytest	7.4.0	MIT License		
15	pkg:pypi/python-dotenv@1.0.0	python-dotenv	1.0.0	BSD License		
16	pkg:pypi/regex@2023.10.3	regex	2023.10.3	Apache Software License		
17	pkg:pypi/reportlab@4.0.7	reportlab	4.0.7	BSD License		
18	pkg:pypi/scikit-learn@1.3.2	scikit-learn	1.3.2	BSD License		

Last Name	First Name						
Baumgärtner	**Lucca**			**1.00**	**OK**		
Bozkurt	**Resit Berkay**	1					
Hambauer	**Nico**						
Heasman	**Sophie**	1					
Nathani	**Ruchita**	1		0	No size		
Sheta	**Ahmed**			1	Trivial size		
Utech	**Fabian-Paul**	1		2	Small size		
Zailskas	**Felix**			3	Medium size		
Zimmermann	**Simon**			5	Large size		
				8	Very large size		
				13	Too large (size)		