Project Name	Sales Lead Qualifier
Online team meeting	https://fau.zoom-x.de/j/61899002491?pwd=WGFIL2tSS2lVWFhqZG9PWmMzVFg4dz09
Production system (if any)	(none)
Test system (if any)	(none)
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	Deliverable / Irr. Work
0	2023-10-18	Simon, Tetiana	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	N/A	Nico		
1	2023-10-25	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Lucca	Nico		X
2	2023-11-08	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Berkay	Nico		X
3	2023-11-15	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ruchita	Nico		
4	2023-11-22	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ahmed	Nico		
5	2023-11-29	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Fabian-Paul	Nico		X
6	2023-12-06	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Felix	Nico	Mid-term due	X
7	2023-12-13	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Sophie	Nico		
8	2023-12-20	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Lucca	Nico		
9	2024-01-10	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Berkay	Nico		
10	2024-01-17	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Felix	Nico		
11	2024-01-24	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Fabian-Paul	Nico		
12	2024-01-31	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Sophie	Nico		X
13	2024-02-07	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ahmed	Nico	Demo day!	X
14	2024-02-14	Simon	Lucca, Berkay, Ruchita, Ahmed, Fabian-Paul, Felix, Sophie	Ruchita	Nico	Retrospective	X

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
Lead	Potential customer of our industry partner which has already declared their interest in purchasing products or services.
Software Component	Major part of our product, which is mostly independent from other components (i.e. can be run separately).
Base Data Collector (BDC)	One of our main two software components which fulfills the task of collecting data about a lead from various online sources.
Merchant Size Predictor (MSP)	Other main software component which performs the task of predicting the sizes of leads. Uses an underlying AI-based predictor.
Controller	Planned software component that orchestrates BDC and MSP, leading them to an efficient collaboration. Not yet implemented.
AI-based predictor	Machine learning model that can be trained on a training data set to make predictions on a data set of similar shape afterwards.
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.
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).
Data Feature	A common property among multiple data records (e.g. first / last name, phone number).
Data Label	A data feature which value is unknown but wich will be calculated by an AI-based predictor.
Provided Data	Data from the internal data source, that contains few data features and is possibly irregular. Input for BDC.
Historical Data	Provided data about previous leads that have become customers.
Lead Data	Provided data about current leads.
Collected Data	All data about a lead or customer, which has been collected from external data source by the BDC.
Enhanced Data	Combination of provided and collected data, possibly enriched by some derived features. Output of BDC.
Preprocessed Data	Data suitable for the usage in an AI-based predictor, either for training or for prediction.
Training Data	Preprocessed enhanced historical data. Input for training an AI-based predictor insode of the MSP.
Prediction Input	Preprocessed enhanced lead data. Input for running an AI-based predictor inside of the MSP.
Prediction Result	Provided data, amended by data labels that were calculated by an AI-based predictor (merchant size in our case). Output of the MSP.

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	Improve the AI model by making it preciser and more computationally efficient.
12	Finalize work on our AI model (fine tuning). Review our current project state to address technical debts.
13	Finalize documentation and polish code base to prepare the handover.
14	Create project report & project retrospective

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			70	70		
Sprints						
7	**Documentation**		12	70	7	70
8	**More Data**		8	58	10	63
9	**Data Preprocessing**		17	50	12	53
10	**Train AI Model**		6	33	5	41
11	**Improve Model**		7	27	7	36
12	**Fine Tune Model**		10	20	10	29
13	**Final Polishing**		21	10	21	19
	Sum			-11		-2
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 Preprocessing**	Possibility to amend to data	5		5	
		Use default pipeline during sprint release	1		1	
		Configure pipeline using a config file	2		0	
		Preprocessing of collected lead data	5		3	
		Bug: Incorporating Regional Atlas doesn't let the pipeline run till the end	1		1	
		Inspect lead "neighbors"	3		2	

Sprint	Goal	Feature Name	Est. Size	Est. Remaining	Real Size	Real Remaining
10	**Train AI Model**	Train ML model prototype with historical data	3		3	
		Systematic approach to find more lead data (once again)	3		2	
11	**Improve Model**	Create improved ML model iteration	5		5	
		Business type analysis	2		2	
12	**Fine Tune Model**	Final iteration of ML model	5		5	
		Investigate our testing setup	3		3	
		Integrate implementation results from #197 & document it	2		2	
13	**Final Polishing**	Simplify Reports	2		2	
		Repair BDC pipeline runs with forceRefresh=False	3		3	
		Remove unused features / dead code	2		2	
		Automatic generation of documentation (pydoc)	3		3	
		Distininguish between lead_preprocessed_data and historical_preprocessed_data	3		3	
		Finalize documentation	2		2	
		Ensure that our project is easy to use	3		3	
		Better Testing	3		3	

#	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
GitHub Pages (Sphinx)	https://amosproj.github.io/amos2023ws06-sales-lead-qualifier/
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	Python Package Index	aiobotocore	2.11.2	Apache Software License	
2	Python Package Index	aiohttp	3.9.3	Apache Software License	
3	Python Package Index	aioitertools	0.11.0	MIT License	
4	Python Package Index	aiosignal	1.3.1	Apache Software License	
5	Python Package Index	annotated-types	0.6.0	MIT License	
6	Python Package Index	anyio	3.7.1	MIT License	
7	Python Package Index	async-timeout	4.0.3	Apache Software License	
8	Python Package Index	attrs	23.2.0	MIT License	
9	Python Package Index	autocorrect	2.6.1	GNU Lesser General Public License v3 (LGPLv3)	
10	Python Package Index	beautifulsoup4	4.12.2	MIT License	
11	Python Package Index	boto3	1.33.1	Apache Software License	
12	Python Package Index	botocore	1.33.13	Apache Software License	
13	Python Package Index	certifi	2024.2.2	Mozilla Public License 2.0 (MPL 2.0)	
14	Python Package Index	charset-normalizer	3.3.2	MIT License	
15	Python Package Index	click	8.1.7	BSD License	
16	Python Package Index	click-plugins	1.1.1	BSD License	
17	Python Package Index	cligj	0.7.2	BSD	
18	Python Package Index	colorama	0.4.6	BSD License	
19	Python Package Index	coloredlogs	15.0.1	MIT License	
20	Python Package Index	ConfigArgParse	1.7	MIT License	
21	Python Package Index	dateparser	1.2.0	BSD License	
22	Python Package Index	de-autobahn	1.0.4	Error	
23	Python Package Index	de-bundesrat	0.1.0	Apache Software License	
24	Python Package Index	de-bundestag	0.1.0	Apache Software License	
25	Python Package Index	de-dwd	1.0.1	Error	
26	Python Package Index	de-interpol	0.1.0	Apache Software License	
27	Python Package Index	de-jobsuche	0.1.0	Apache Software License	
28	Python Package Index	de-ladestationen	1.0.5	Apache Software License	
29	Python Package Index	de-mudab	0.1.0	Apache Software License	
30	Python Package Index	de-nina	1.1.0	Apache Software License	
31	Python Package Index	de-polizei-brandenburg	0.1.0	Apache Software License	
32	Python Package Index	de-risikogebiete	0.1.0	Apache Software License	
33	Python Package Index	de-smard	0.1.0	Apache Software License	
34	Python Package Index	de-strahlenschutz	1.0.0	Apache Software License	
35	Python Package Index	de-travelwarning	0.1.0	Apache Software License	
36	Python Package Index	de-zoll	0.1.0	Apache Software License	
37	Python Package Index	deep-translator	1.11.4	MIT License	
38	Python Package Index	deutschland	0.4.0	Apache Software License	
39	Python Package Index	distro	1.9.0	Apache Software License	
40	Python Package Index	dnspython	2.5.0	ISC License (ISCL)	
41	Python Package Index	email-validator	2.1.0.post1	CC0 1.0 Universal (CC0 1.0) Public Domain Dedication	

#	Context	Name	Version	License	Comment
42	Python Package Index	exceptiongroup	1.2.0	MIT License	
43	Python Package Index	filelock	3.13.1	The Unlicense (Unlicense)	
44	Python Package Index	fiona	1.9.5	BSD License	
45	Python Package Index	flatbuffers	23.5.26	Apache Software License	
46	Python Package Index	frozenlist	1.4.1	Apache Software License	
47	Python Package Index	fsspec	2023.12.2	BSD License	
48	Python Package Index	geopandas	0.14.1	BSD License	
49	Python Package Index	googlemaps	4.10.0	Apache Software License	
50	Python Package Index	h11	0.14.0	MIT License	
51	Python Package Index	httpcore	1.0.2	BSD License	
52	Python Package Index	httpx	0.26.0	BSD License	
53	Python Package Index	humanfriendly	10.0	MIT License	
54	Python Package Index	idna	3.6	BSD License	
55	Python Package Index	Jinja2	3.1.3	BSD License	
56	Python Package Index	jmespath	1.0.1	MIT License	
57	Python Package Index	joblib	1.3.2	BSD License	
58	Python Package Index	lightgbm	4.3.0	MIT License	
59	Python Package Index	lxml	4.9.4	BSD License	
60	Python Package Index	mapbox-vector-tile	2.0.1	MIT License	
61	Python Package Index	MarkupSafe	2.1.5	BSD License	
62	Python Package Index	more-itertools	8.14.0	MIT License	
63	Python Package Index	mpmath	1.3.0	BSD License	
64	Python Package Index	multidict	6.0.5	Apache Software License	
65	Python Package Index	networkx	3.2.1	BSD License	
66	Python Package Index	nltk	3.8.1	Apache Software License	
67	Python Package Index	numpy	1.26.1	BSD License	
68	Python Package Index	onnxruntime	1.17.0	MIT License	
69	Python Package Index	openai	1.3.3	Error	
70	Python Package Index	osmnx	1.7.1	MIT License	
71	Python Package Index	packaging	23.2	Apache Software License, BSD License	
72	Python Package Index	pandas	2.0.3	BSD License	
73	Python Package Index	phonenumbers	8.13.25	Apache Software License	
74	Python Package Index	pillow	10.2.0	Historical Permission Notice and Disclaimer (HPND)	
75	Python Package Index	protobuf	4.25.2	3-Clause BSD License	
76	Python Package Index	pyclipper	1.3.0.post5	MIT License	
77	Python Package Index	pydantic	2.6.0	MIT License	
78	Python Package Index	pydantic_core	2.16.1	MIT License	
79	Python Package Index	pyLanguagetool	0.10.0	MIT License	
80	Python Package Index	pyproj	3.6.1	MIT License	
81	Python Package Index	pyspellchecker	0.7.2	MIT License	
82	Python Package Index	python-dateutil	2.8.2	Apache Software License, BSD License	

#	Context	Name	Version	License	Comment
83	Python Package Index	python-dotenv	0.21.0	BSD License	
84	Python Package Index	pytz	2024.1	MIT License	
85	Python Package Index	regex	2023.12.25	Apache Software License	
86	Python Package Index	reportlab	4.0.7	BSD License	
87	Python Package Index	requests	2.31.0	Apache Software License	
88	Python Package Index	s3fs	2023.12.2	BSD License	
89	Python Package Index	s3transfer	0.8.2	Apache Software License	
90	Python Package Index	scikit-learn	1.3.2	BSD License	
91	Python Package Index	scipy	1.12.0	BSD License	
92	Python Package Index	shapely	2.0.2	BSD License	
93	Python Package Index	six	1.16.0	MIT License	
94	Python Package Index	sniffio	1.3.0	Apache Software License, MIT License	
95	Python Package Index	soupsieve	2.5	MIT License	
96	Python Package Index	sympy	1.12	BSD License	
97	Python Package Index	textblob	0.17.1	MIT License	
98	Python Package Index	threadpoolctl	3.2.0	BSD License	
99	Python Package Index	tiktoken	0.5.1	MIT License	
100	Python Package Index	torch	2.1.2	BSD License	
101	Python Package Index	tqdm	4.65.0	MIT License, Mozilla Public License 2.0 (MPL 2.0)	
102	Python Package Index	typing_extensions	4.9.0	Python Software Foundation License	
103	Python Package Index	tzdata	2023.4	Apache Software License	
104	Python Package Index	tzlocal	5.2	MIT License	
105	Python Package Index	urllib3	2.0.7	MIT License	
106	Python Package Index	wrapt	1.16.0	BSD License	
107	Python Package Index	xgboost	2.0.3	Apache Software License	
108	Python Package Index	yarl	1.9.4	Apache Software License	

Last Name	First Name				
Baumgärtner	**Lucca**			**3.00**	**OK**
Bozkurt	**Resit Berkay**	3			
Hambauer	**Nico**				
Heasman	**Sophie**	3			
Nathani	**Ruchita**	3		0	No size
Sheta	**Ahmed**	3		1	Trivial size
Utech	**Fabian-Paul**			2	Small size
Zailskas	**Felix**	3		3	Medium size
Zimmermann	**Simon**			5	Large size
				8	Very large size
				13	Too large (size)