Lessons Learned From Software Maintenance and Support Datasets

Arlene Minkiewicz
Chief Scientist
PRICE Systems LLC
Arlene.Minkiewicz@pricesystems.com

Estimate with Confidence™

PRICE®

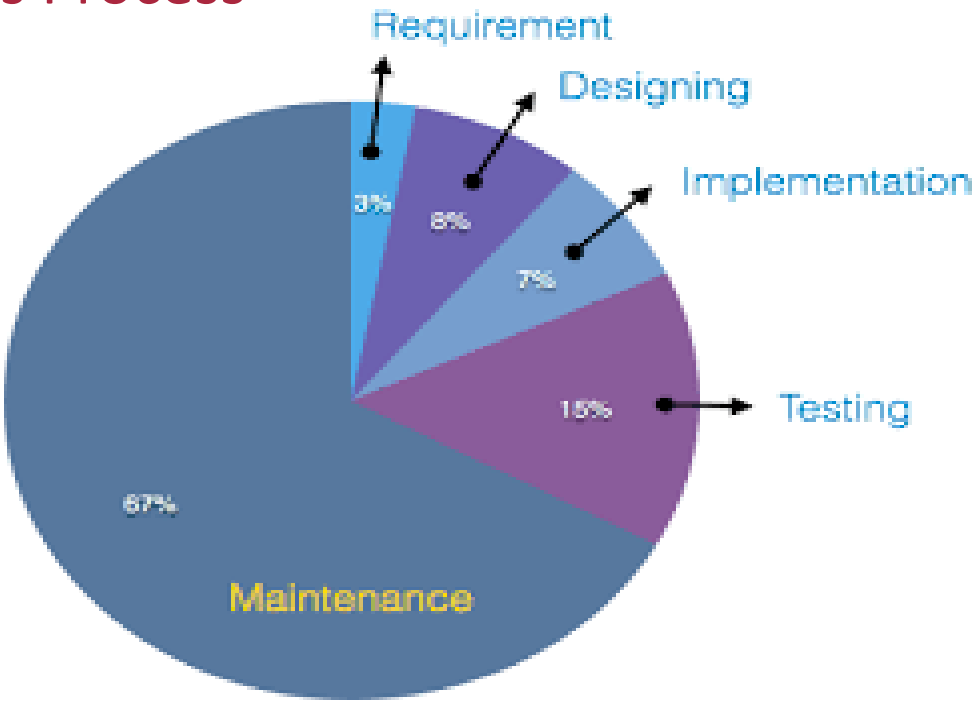
Cost Analytics

TrueExplorer	TrueFindings	PRICE® Models	TruePlanner	TrueMapper	TrueBOE	TrueXLS
Search & extract data from the PCA Ecosystem	Manage & Analyze Data Sets	Predictive Models	Integration Framework	Customer Data Mapping	Basis-of-Estimate Generator	Access PCA Engine from Excel

Excel Word Project @RISK ORACLE CRYSTAL BALL PROPRICER PHOENIX INTEGRATION ACEIT

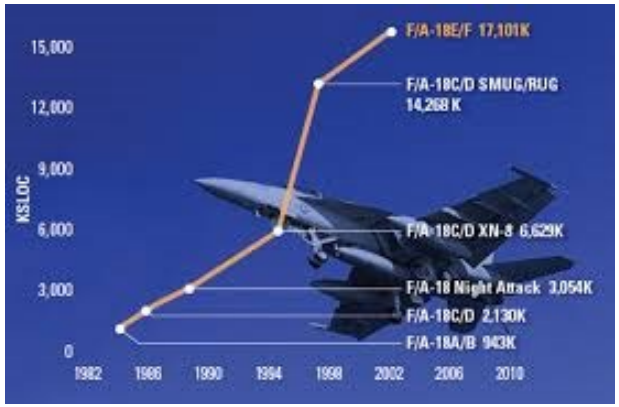
Agenda

- Introduction
  - Software Sustainment
  - ISBSG
- Data Collection and Analysis Process
- ISBSG Data Study
  - Study Objectives
  - Data Collection and Preparation
  - Modeling and Evaluation
  - Deployment
- Lessons Learned
- Future Work



Introduction

- Software Sustainment has become an increasingly important part of the operation of complex systems
  - Software is more malleable than hardware, easier to change 'on the fly'
  - The life of many software projects have greatly exceeded their expected lifetime
  - Changes need to be made to the software to evolve it as hardware and software technology evolves
  - Software is easier to deploy than hardware – sometimes used to correct hardware inadequacies
- While many studies have been done over time on the estimation of software sustainment costs, few have much publicly available data behind them
- This study presents lessons learned from the ISBSG database

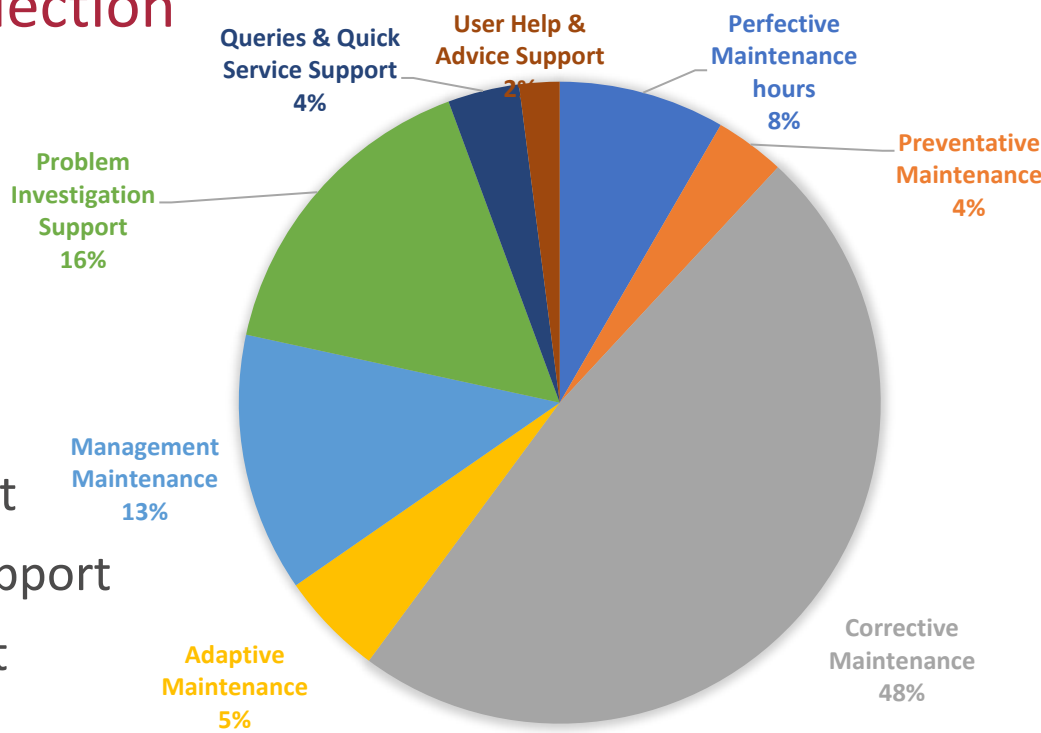


Software Sustainment

- Many definitions for software sustainment
- *"Software sustainment involves orchestrating the processes, practices, technical resources, information and workforce competencies for systems and software engineering, to enable system to continue mission operations and also to be able to be enhanced to meet evolving threat and capability needs."*
  - Software Engineering Institute at Carnegie Mellon University (SEI/CMU)
- *Software maintenance is "the process of modifying a software system after delivery to correct faults, improve performance and adapt to changing environments."*
  - Institute of Electrical and Electronics Engineering's (IEEE) Standard 12207
- Software Sustainment is generally considered to be all activities associated with a software application after it is first released to the public

Software Sustainment

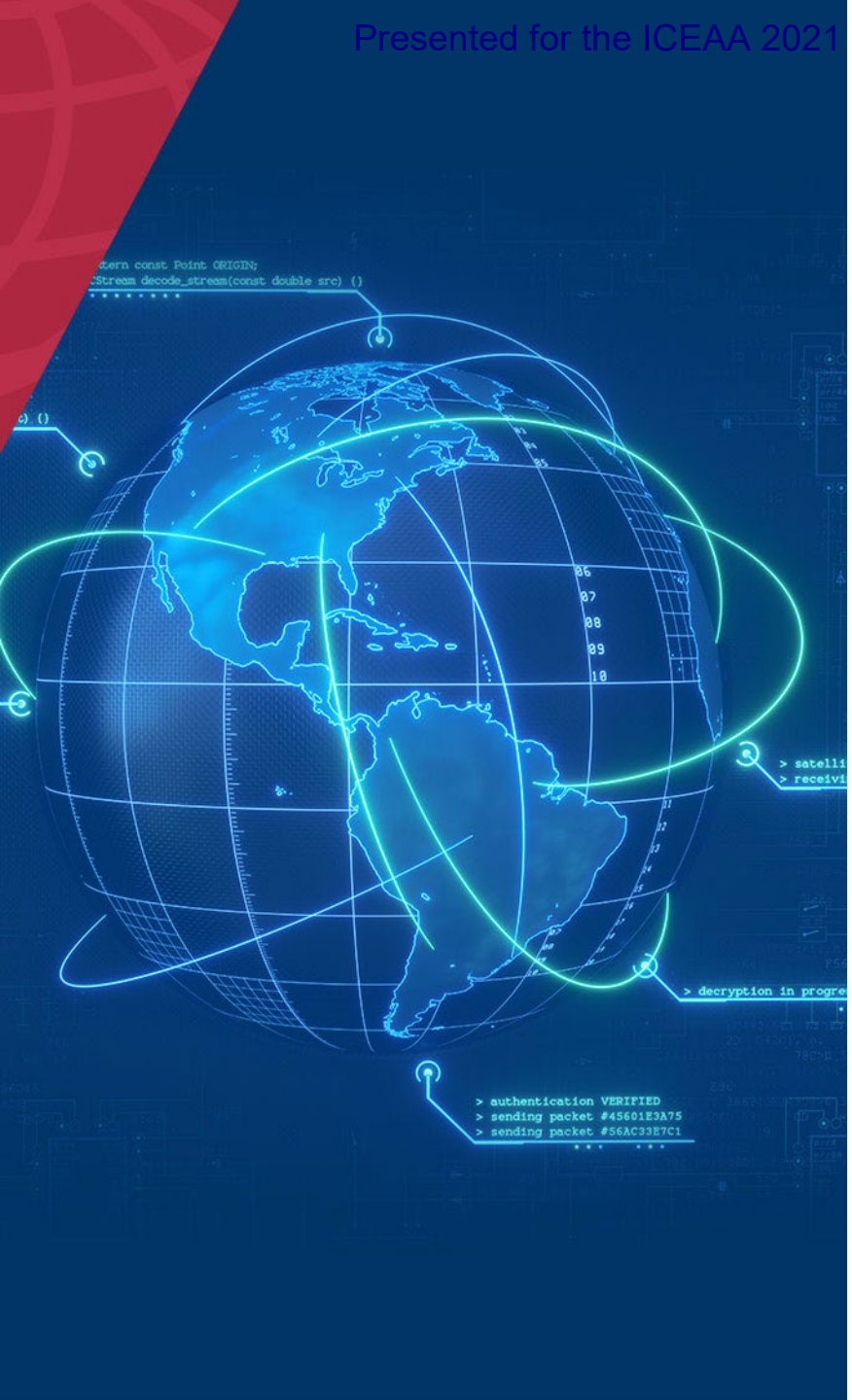
- For the purposes of this study, Software Sustainment will encompass the activities outline in the ISBSG Attributes associated with effort collection
  - Adaptive Maintenance
  - Corrective Maintenance
  - Perfective Maintenance
  - Preventative Maintenance
  - Maintenance Management
  - Problem Investigation Support
  - Queries and Quick Service Support
  - User Help and Advice Support
  - Support Management



ISBSG

- International Software Benchmark Standards Group
  - Not for profit
  - Founded in Australia in 1997
  - Mission to help the Information Technology (IT) industry improve software processes and Practices
- All ISBSG Data is:
  - Validated and rated in accordance with the ISBSG Quality Guidelines
  - Representative of the industry
  - Independent and Trusted
  - Captured from a range of organizations and industries
  - Anonymously stored in the repository
- In August 2020 Version 7 of the ISBSG Maintenance and Support (M&S) database was release
- Study data was collected from this database

www.isbsg.org | ISBSG Delivering IT Confidence



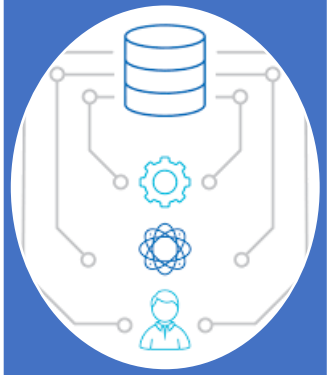
Data Collection and Analysis

Data Collection and Analysis Process

- Data collection is a daunting task
- Data mining is the process of finding anomalies, patterns and correlations withing data sets to predict outcomes and to turn raw data into useful information.
- Sound data mining practices applied sensibly are an important success criteria for data collection and analysis projects

Data Collection and Analysis Process

Understand the data challenge



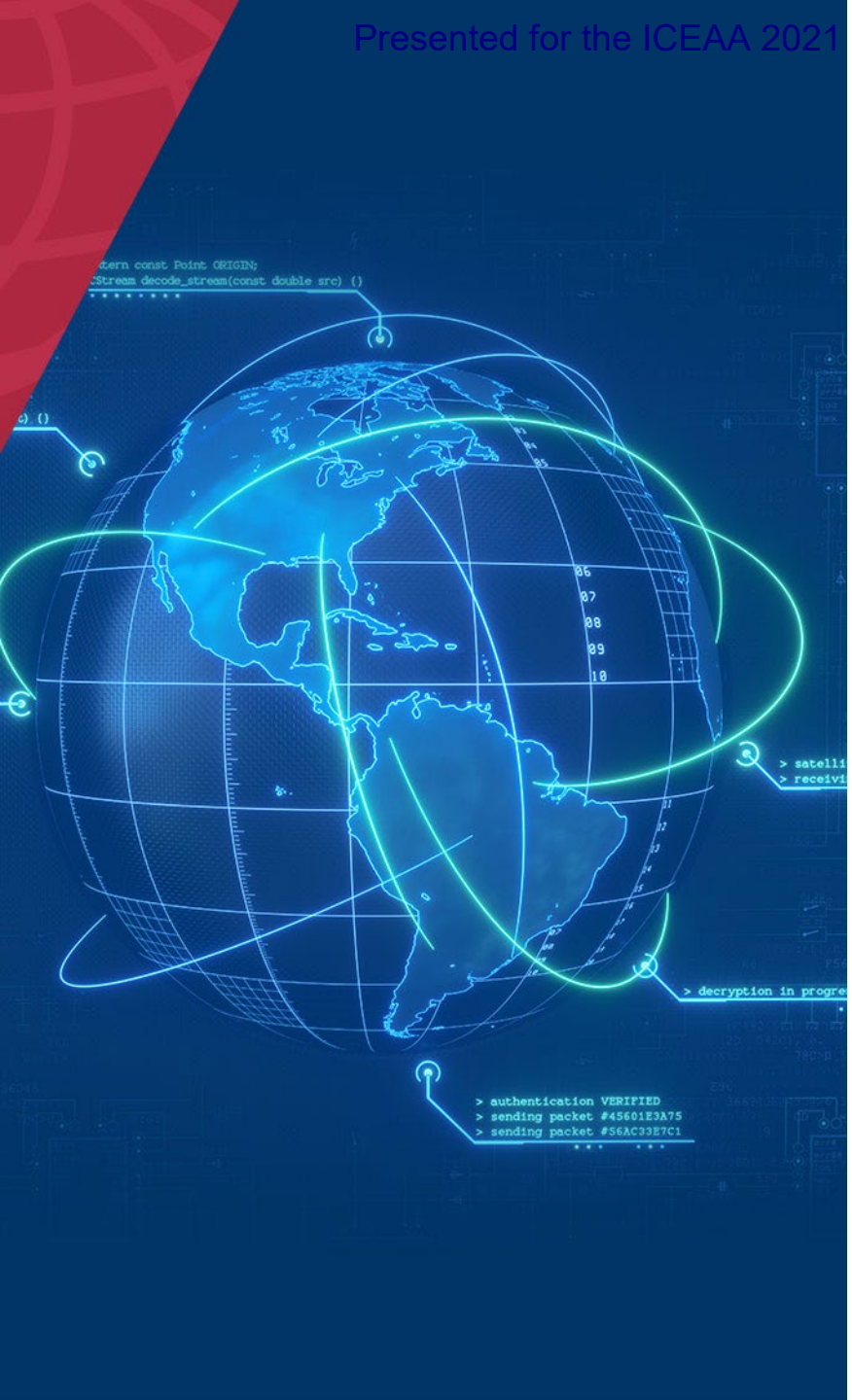
Data Understanding and Preparation

Modeling

Evaluation

Deployment

Iterative Process



ISBSG Data Study

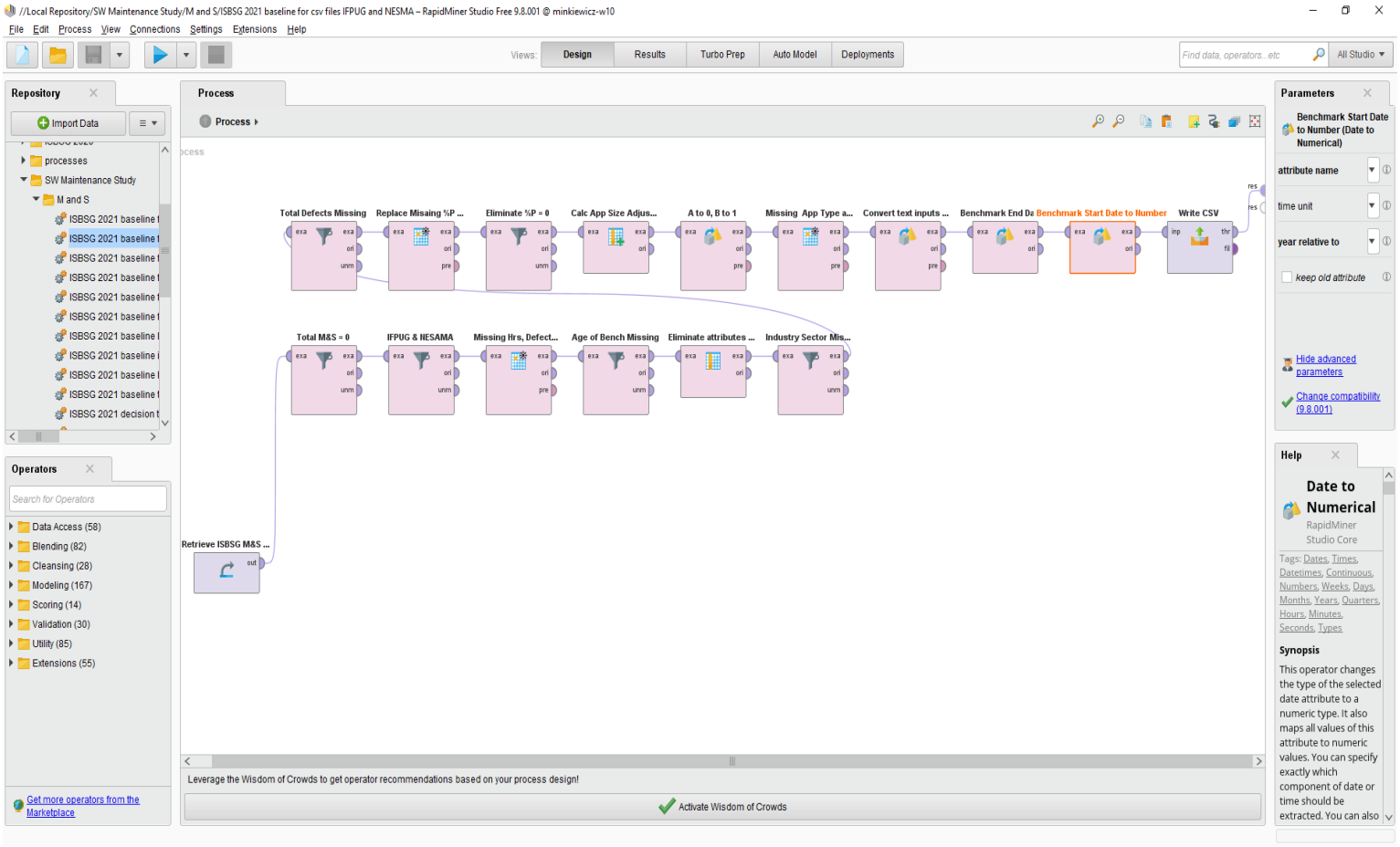
The Questions To Be Answered

- What lessons can be learned from the ISBSG M&S Data?
- How can these lessons be used to provide software maintenance and support estimation guidance to the cost community?
- What useful benchmarks can be gleaned from the industries to support estimation efforts?
- How do Source Lines of Code compare to Functional Size Measurements with respect to effort and defect estimation?

Data Preparation and Understanding

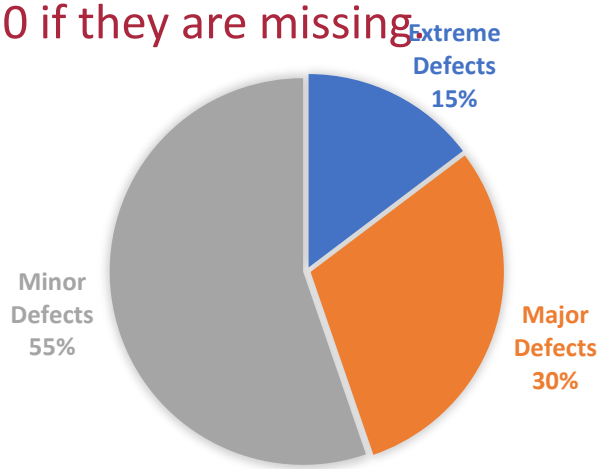
- ISBSG M&S Database
  - 1177 Data Points
  - 114 potential features (attributes)
- Two datasets to be collected
  - Application Sizing Approach is either IFPUG or NESMA
  - Entered KSLOC is greater than 0
- Acquire the right tools for the next steps
  - RapidMiner
  - TrueFindings®
  - Microsoft® Excel

Data Preparation and Understanding



Data Preparation and Understanding

- Operators in the RM Process accomplish the following tasks
- Retrieve the ISBSG M&S Data
  - Since the data from the ISBSG was in the Excel format it was easily imported into RapidMiner for use in this process.
- Filter out all records for which Total Maintenance and Support Hours = 0.
- Filter to develop the two Datasets
  - Application Size Approach = IFPUG or NESMA
  - Entered KSLOC > 0
- For the records that remain set the following values to 0 if they are missing
  - Adaptive Maintenance Hours
  - Corrective Maintenance Hours
  - Perfective Maintenance Hours
  - Preventative Maintenance Hours
  - Problem Investigation Support Hours
  - Queries and Quick Service Support Hours
  - User Help and Advice Support
  - Minor Defects
  - Major Defects
  - Extreme Defects
- Removed records where the Age of Benchmark attribute was missing.



Data Preparation and Understanding

- Operators in the RM Process accomplish the following tasks (cont.)

- Eliminate attributes that are not needed for this study or are scarcely populated. Appendix C contains a list of the attributes eliminated along with the rationale.
- Filter out records missing Industry Sector.
- Eliminate records with Total Defects Missing
- Replace Missing Percent %P with 100 – %P is the percent of Primary Language in the application.
- Eliminate records where the %P is equal to 0 as this makes the data record suspicious.
- Convert Data Quality Factors to Numerical values to make numerical analyses possible.
- Replace Application Type and Primary Language with Numerical values to make numerical analyses possible.
- Replace Benchmark Start and End Date with Numerical Values
- Write resulting example set to a CSV file for importation into TrueFindings®.

Benchmarks for each Data Set

- Benchmarks across both data sets (IFPUG and KSLOC) were compiled for:
  - Application Size
  - Hours (Adaptive, Corrective, Preventive, Perfective, Management, Problem Investigation, Queries and Quick Service, User Help and Advice, Maintenance, Support and Total M&S)
  - Hours Per Size Unit for each type of Hours mentioned above
  - Defect counts (Minor, Major & Extreme)
- Benchmarks include:
  - Sample Size
  - Min
  - Max
  - 25%
  - 75%
  - Median
  - Mode
  - Standard Deviation
  - Coefficient of Variation

Sample Benchmark Data

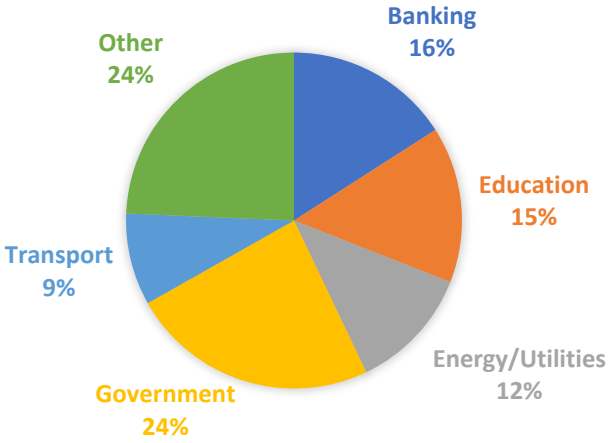
Attribute	Sample Size	Min	Max	25%	75%	Mean	Median	Mode	Standard Deviation	Coefficient of Variation
Age at Benchmark	226	1	31	4	12	8.85	8	3	6.018	0.68
Adaptive Hours Per FP	30	0.002	3.894	0.055	0.423	0.44	0.197	0.002	0.768	1.746
Adaptive Hours	30	6	1032	21.75	388.75	235	69	18	304.77	1.297
Application Size	226	24	55312	459	3477.8	2931.24	1261.5	234	4893.831	1.67
Corrective FP Rate	41	0.002	4.292	0.038	0.327	0.356	0.095	0.002	0.742	2.083
Corrective Maintenance Hours	41	2	4224	24	132	320.707	48	6	843.871	2.631
Extreme Defects per FP	183	1.43E-04	0.066	0.002	0.006	0.005	0.004	0.002	0.005	1.12
Extreme Defects	183	1	152	2	15.5	12.778	5	2	18.525	1.45
Maintenance FP Rate	208	0.1	6.3	0.2	0.411	0.414	0.307	0.2	0.624	1.506
Major Defect FP Rate	185	3.10E-04	1	0.006	0.012	0.024	0.008	0.008	0.111	4.58
Major Defects	185	1	326.51	5	32	27.4	14	3	41.697	1.522
Management Maintenance Hours	172	2	3139	66.25	283.5	215.326	149	52	291.266	1.353
Minor Defects Per FP	193	3.10E-04	0.667	0.016	0.036	0.037	0.022	0.05	0.065	1.729
Minor Defects	193	1	954	19	82	69.658	41	12	121.581	1.745
Perfective Hours per FP	26	0.006	0.43	0.03	0.094	0.078	0.065	0.086	0.084	1.076
Preventive Maintenance Hours	34	1	240	6.75	76.5	60.853	31	6	71.711	1.178
Preventive FP Rate	34	0.002	0.747	0.034	0.113	0.116	0.068	0.1	0.16	1.377
Problem Investiageion FP Rate	31	0.003	2.113	0.041	0.234	0.226	0.098	0.003	0.408	1.802
Problem Imvestigation Hours	31	8	1237	23.5	106.5	112.258	60	13	220.839	1.967
Queries & Quick Service Support Hours	33	1	3752	13	132	243.909	67	13	656.753	2.693
Queries and Quick Service FP Rate	33	0.003	1.1	0.055	0.271	0.222	0.12	0.003	0.274	1.233
Support FP Rate	204	0.1	3.3	0.186	0.378	0.339	0.3	0.2	0.337	0.992
total Defects FP Rate	219	0.001	1.667	0.023	0.05	0.059	0.034	0.043	0.157	2.647
Total Defects	219	1	1311	22.5	111.5	96.831	51	32	159.462	1.647
Total Maintenance and Support Hours	226	11	22423	329.3	1759	1511.04	881.5	52	2457.856	1.627
Total Maintenance Hours	219	6	13182	181	936.5	780.174	452	126	1327.703	1.702
Total Support Hours	208	8	11212	170.8	841.25	687.611	453.5	40	1021.549	1.486
User Help and Advise Support Hours	27	3	495	12.5	99.5	72.074	33	13	100.605	1.396
User Help and Advice FP Rate	27	0.004	1.1	0.027	0.147	0.128	0.051	0.004	0.22	1.72

Additional Benchmark Scenarios

- IFPUG/NESMA Dataset
  - Primary Language
    - *.Net*
    - *COBOL*
    - *Java*
    - *Oracle*
  - Application Type
    - *Business Applications*
    - *CRM/ERP*
    - *Document Management*
    - *Web Application*
  - Industry Sector
    - *Education*
    - *Financial*
    - *Government*
    - *Logistics*
    - *Utilities*
  - Organization Type
    - *Banking*
    - *Transport*

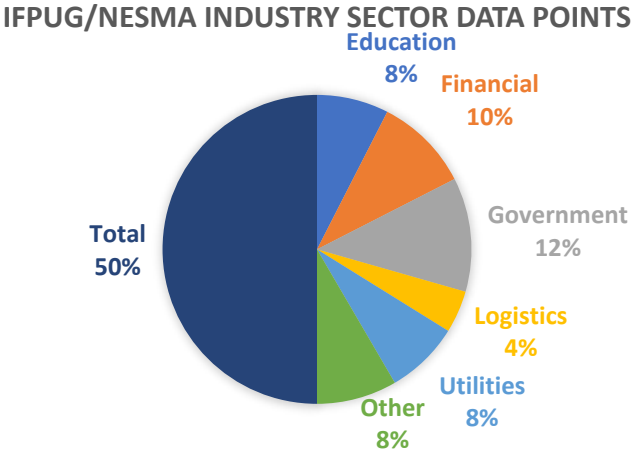
- KSLOC Dataset
  - Primary Language
    - *COBOL*
    - *Java*
    - *JCL*
    - *Telon*
  - Industry Sector
    - *Banking*
    - *Education*
    - *Government*
    - *Wholesale and Retail*

IFPUG/NESMA ORGANIZATION TYPE DATA POINTS



Sample Benchmark Data

Government										
Attribute	Sample Size	Min	Max	25%	75%	Mean	Median	Mode	Standard Deviation	Coefficient of Variation
Age at Benchmark	54	1	31	6	12	9.981	9	6	6.299	0.631
Application Size	54	24	18500	444.75	3542	2724.5	871.5	453	3578.376	1.313
Maintenance FP Rate	45	0.1	2.1	0.206	0.374	0.383	0.287	0.2	0.357	0.931
Support FP Rate	38	0.112	0.732	0.197	0.363	0.307	0.29	0.112	0.142	0.462
Total Defects FP Rate	54	0.003	1.667	0.022	0.071	0.122	0.037	0.043	0.305	2.499
Total Defects	54	5.415	1037	28.25	121.3	101.194	48.5	32	155.578	1.537
Total Maintenance & Support Hours	54	13	15338	273.5	1824	1515.13	809	13	2385.777	1.575
Total Maintenance Hours	52	13	4657	137.5	941	718.635	364.5	225	926.848	1.29
Total Maintenance & Support FP Rate	54	0.038	4.75	0.367	0.759	0.728	0.606	0.038	0.732	1.005
Total Support Hours	38	91	3153	235.25	967.5	734.579	495.5	91	662.119	0.901



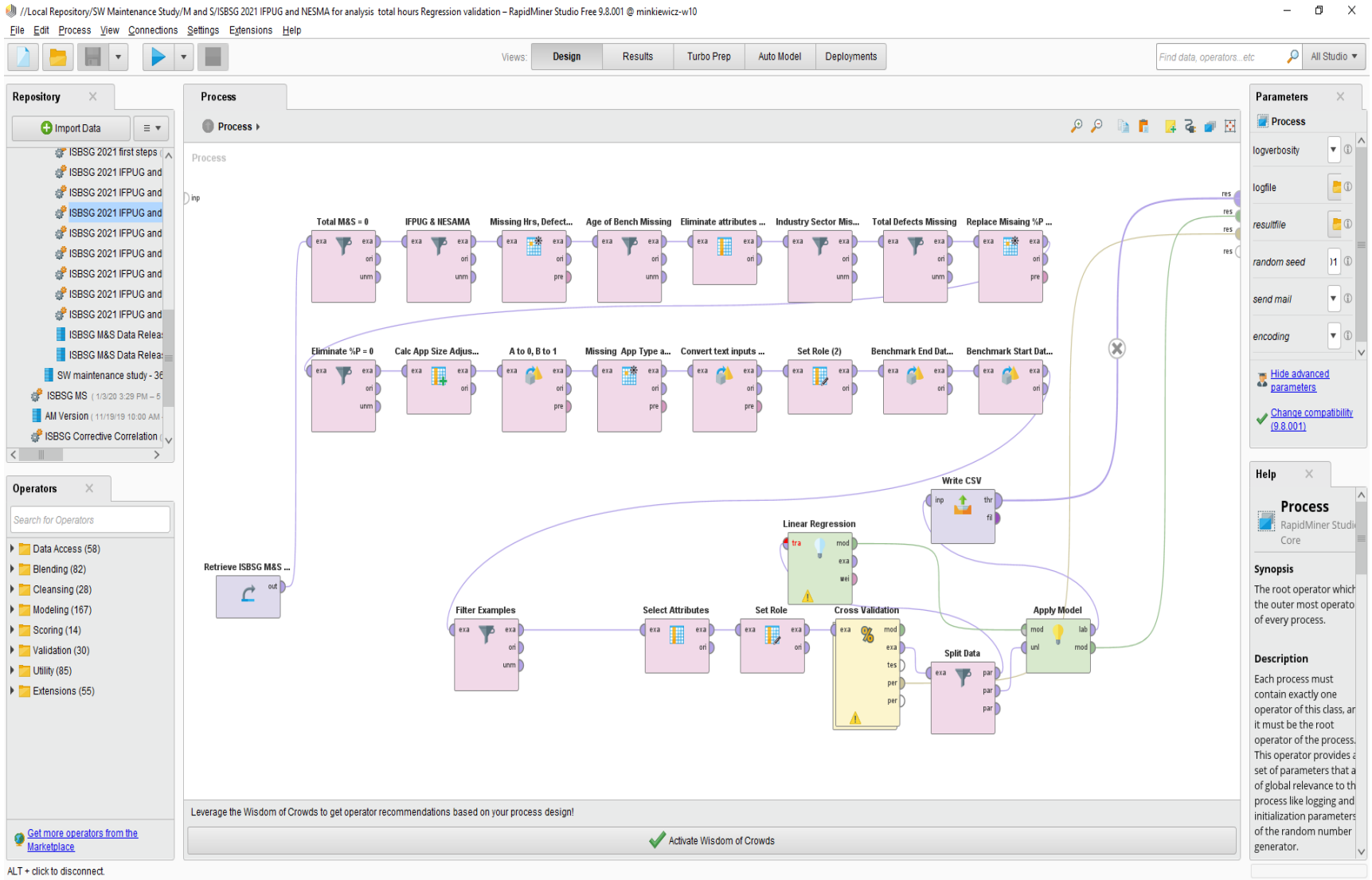
Modeling and Evaluation

- Visualization and Correlation Analysis rendered several initial conclusions
  - There was not very much interesting to learn from the KSLOC data set
    - *Possibly due to the primary focus of ISBSG Data collection – which is Functional Size*
    - *Further analysis on the KSLOC data set was postponed to a future time when more data is available*
  - There appeared to be inconsistencies between the Total Hours for Support and Maintenance as compared to the categorial hours
  - The relationships around Size and Defects told a more compelling story about Support than Maintenance
- Modeling was performed using RapidMiner. Regression Analysis was the primary technique

Modeling and Evaluation

Attributes	Total M...	Perfecti...	Prevent...	Correcti...	Adaptiv...	Manage...	Total Su...	Proble...	Queries...	User He...	Manage...	Extrem...	Major D...	Minor D...	Age at ...	Total M...	Applicat...	Total De...	CalcAp...
Total Maintenance hours	1	0.140	0.152	0.253	0.045	0.661	0.810	0.129	0.002	0.101	-0.033	0.538	0.383	0.457	0.144	0.835	0.708	0.495	0.708
Perfective Maintenance hours	0.140	1	0.782	0.555	0.454	-0.140	-0.025	0.441	0.198	0.392	0.164	-0.115	-0.113	-0.009	-0.033	0.054	0.053	-0.037	0.052
Preventative Maintenance hours	0.152	0.782	1	0.625	0.522	-0.159	-0.034	0.550	0.175	0.513	-0.025	-0.138	-0.133	-0.024	-0.103	0.054	0.026	-0.055	0.027
Corrective Maintenance hours	0.253	0.555	0.625	1	0.102	-0.094	0.023	0.577	0.152	0.501	0.005	-0.057	-0.044	0.129	-0.027	0.137	0.144	0.079	0.142
Adaptive Maintenance hours	0.045	0.454	0.522	0.102	1	-0.140	-0.058	0.217	0.155	0.351	0.075	-0.132	-0.131	-0.116	-0.091	-0.011	-0.071	-0.112	-0.069
Management Maintenance hours	0.661	-0.140	-0.159	-0.094	-0.140	1	0.910	-0.105	-0.083	-0.127	-0.055	0.798	0.586	0.610	0.263	0.692	0.907	0.672	0.914
Total Support hours	0.810	-0.025	-0.034	0.023	-0.058	0.910	1	0.083	0.213	0.053	-0.058	0.768	0.674	0.642	0.230	0.807	0.890	0.720	0.884
Problem Investigation Support hours	0.129	0.441	0.550	0.577	0.217	-0.105	0.083	1	0.219	0.855	-0.005	-0.071	-0.098	-0.058	-0.096	0.094	-0.033	-0.068	-0.034
Queries & Quick Service Support hours	0.002	0.198	0.175	0.152	0.155	-0.083	0.213	0.219	1	0.209	-0.011	0.014	0.452	0.244	0.022	0.081	0.034	0.297	-0.007
User Help & Advice Support hours	0.101	0.392	0.513	0.501	0.351	-0.127	0.053	0.855	0.209	1	-0.001	-0.096	-0.119	-0.080	-0.111	0.065	-0.046	-0.089	-0.046
Management Support Hours	-0.033	0.164	-0.025	0.005	0.075	-0.055	-0.058	-0.005	-0.011	-0.001	1	-0.029	-0.012	-0.038	-0.068	-0.046	-0.024	-0.035	-0.027
Extreme Defects	0.538	-0.115	-0.138	-0.057	-0.132	0.798	0.768	-0.071	0.014	-0.096	-0.029	1	0.759	0.638	0.119	0.572	0.851	0.758	0.851
Major Defects	0.383	-0.113	-0.133	-0.044	-0.131	0.586	0.674	-0.098	0.452	-0.119	-0.012	0.759	1	0.724	0.117	0.509	0.682	0.854	0.656
Minor Defects	0.457	-0.009	-0.024	0.129	-0.116	0.610	0.642	-0.058	0.244	-0.080	-0.038	0.638	0.724	1	0.168	0.712	0.733	0.973	0.703
Age at Benchmark	0.144	-0.033	-0.103	-0.027	-0.091	0.263	0.230	-0.096	0.022	-0.111	-0.068	0.119	0.117	0.168	1	0.194	0.198	0.161	0.195
Total Maintenance & Support hours	0.835	0.054	0.054	0.137	-0.011	0.692	0.807	0.094	0.081	0.065	-0.046	0.572	0.509	0.712	0.194	1	0.752	0.712	0.735
Application size	0.708	0.053	0.026	0.144	-0.071	0.907	0.890	-0.033	0.034	-0.046	-0.024	0.851	0.682	0.733	0.198	0.752	1	0.795	0.998
Total Defects	0.495	-0.037	-0.055	0.079	-0.112	0.672	0.720	-0.068	0.297	-0.089	-0.035	0.758	0.854	0.973	0.161	0.712	0.795	1	0.767
CalcApplicationSize	0.708	0.052	0.027	0.142	-0.069	0.914	0.884	-0.034	-0.007	-0.046	-0.027	0.851	0.656	0.703	0.195	0.735	0.998	0.767	1

Modeling and Evaluation

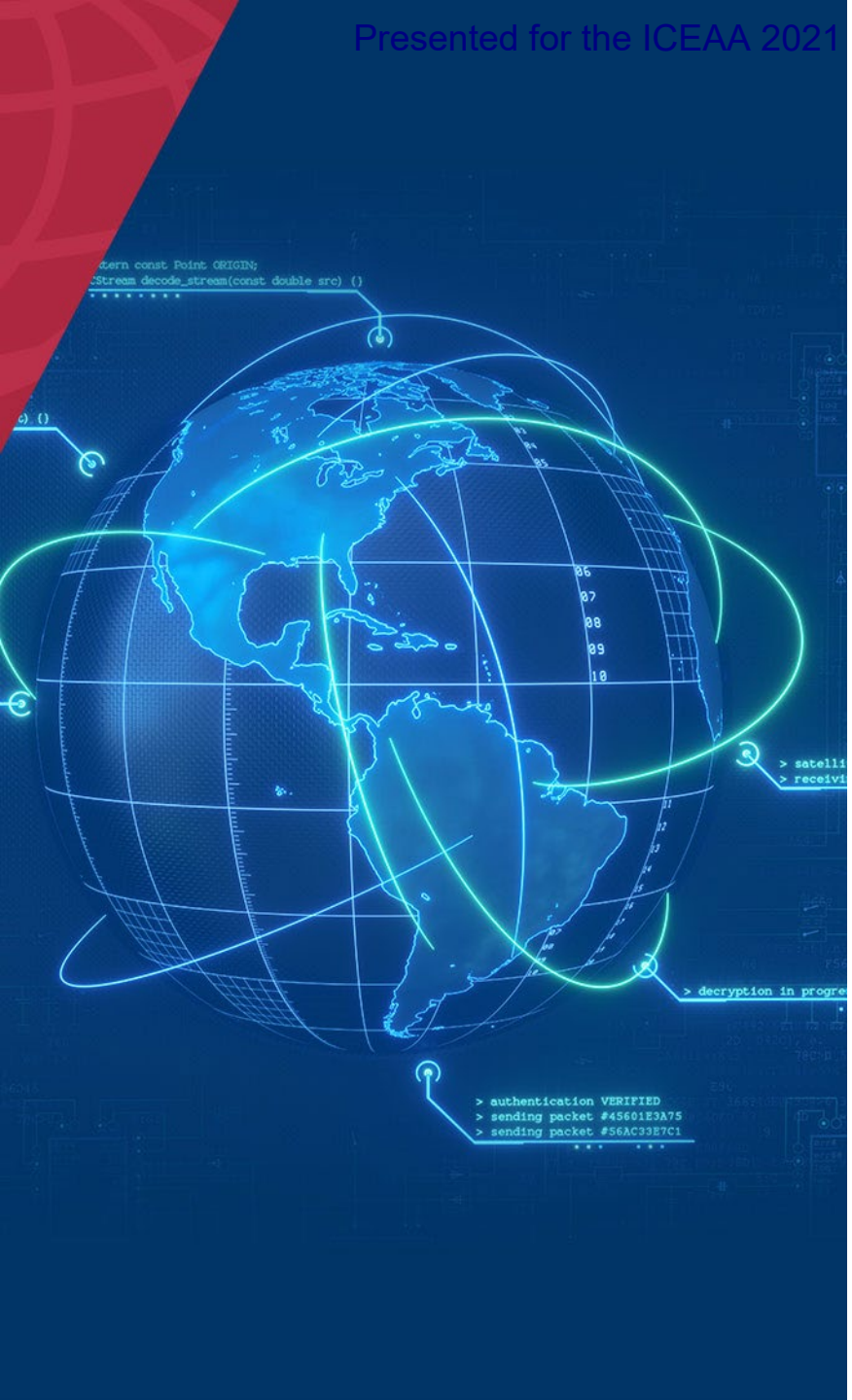


Modeling and Evaluation

- Filter example is used to remove values of 0 from the data set for the attributes that are slated to be independent and dependent variables in the analysis.
  - For example - some examples do not have Support Hours because it was not reported by the submitter (possibly the organization does not track that, or another department handles it) so that data point would not be a valuable addition to a study looking at drivers for Support Hours.
- Select Attributes selects those attributes that represent the dependent and independent variables for a particular analysis.
- Set Role labels the Dependent variable for a particular analysis.
- The Cross Validation Operator performs a Linear Regression on the entire data set and returns performance parameters.
- The Split Data Operator splits the data into a training set (75% of examples) and a scoring set (25% of examples)
- The Model Created by the Linear Regression Operator and the scoring set are sent to the Apply Model which determines how well the model performs on data not in the training set

Modeling and Evaluation

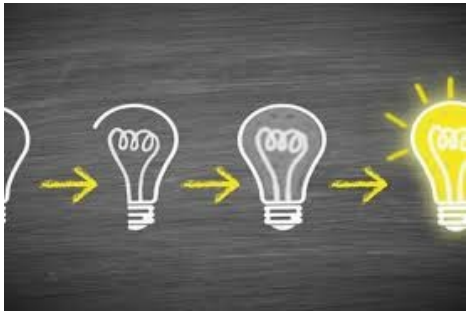
				Training Results			Scoring Results			
Dependent Variable	Independent Variable 1	Independent Variable 2	Independent Variable 3	Equation	R2 (equation)	Training Count	Scoring Count	R2 (prediction)	Pred(30)	Pred(50)
Total Maintenance Hours	Application Size(app)	Extreme Defects (exd)	Major Defects (mad)	165.806+0.2469*app-12.171*exd-2.604*mad	0.949	134	45	0.888	62%	89%
Total Mainenance & Support Hours	Application Size(app)	Extreme Defects (exd)	Major Defects (mad)	226.824+0.3977*app-0.54477*exd-3.6202*mad	0.927	135	45	0.483	58%	76%
Total Support Hours	Total Maintenance Hours(totMaint)			202.89+0.621*totMaint	0.910	152	51	0.959	47%	76%
Total Support Hours	Application Size(app)	Extreme Defects (exd)	Major Defects (mad)	125.929 +0.2366*app-5.015*exd-5.125*mad	0.879	117	39	0.879	60%	88%
Total Support Hours	Application Size(app)			85.746+0.193*app	0.855	164	55	0.528	56%	78%
Extreme Defects	Application Size(app)			2.81+0.0029*app	0.723	137	46	0.792	30%	52%
Total Support Hours	Total Defects (td)			181.055+5.759*td	0.700	151	50	0.571	30%	54%
Total Maintenance Hours	Total Defects (td)			383.085 + 4.828*td	0.660	158	53	0.593	22%	45%
Total Support Hours	Extreme Defects (exd)	Major Defects (mad)		90.07+10.798*exd+20.812*mad	0.659	131	43	0.684	28%	65%
Total Defects	Application Size(app)			314.571+31.084	0.632	164	55	0.643	40%	64%
Support Hours	Major Defects (mad)			89.802+26.7155*mad	0.614	131	43	0.674	27%	60%
Total Maintenance & Support Hours	Application Size(app)			451.95+0.382*app	0.570	170	56	0.796	43%	70%
Total Maintenance Hours	Extreme Defects (exd)	Major Defects (mad)		160.733+52.734*exd-1.757*mad	0.569	134	45	0.388	42%	51%
Total Maintenance Hours	Extreme Defects (exd)			187.36+46.755*exd	0.537	135	45	0.239	22%	56%
Total Support Hours	Extreme Defects (exd)			263.9+32.074*exd	0.528	131	44	0.740	25%	41%
Maintenance Hours	Application Size(app)			285.4212+0.173*app	0.510	164	55	0.988	41%	41%
Total Maintenance & Support Hours	Total Defects (td)			565.86+11.33*td	0.510	164	55	0.825	35%	51%
Major Defects	Application Size(app)			10.00+0.0054*APP	0.446	139	46	0.393	19%	50%



Lessons Learned & Future Works

Lessons Learned to support estimation

- In cases where the domains and project characteristics are similar:
  - Industry specific benchmarks
  - Programming language specific benchmarks
  - Organizational benchmarks
  - Rules of thumb for allocating maintenance and support efforts throughout the lifecycle
  - Rules of thumb for predicting defect types for a project
- The resulting CERs, though not ideal, provide some insight into areas where the things we know about our project might help discover things unknown or only suspected:
  - Potential predictors for Software Support
  - Potential predictors for Software Maintenance
  - Potential predictors for software program Defects



Lessons Learned about SW Data Collection

- Qualitative fields should not be freeform
- Application Set Information would be much more useful if companies and projects could be assigned (anonymously) for cross tracking
- Relationship between Total Hours and Hours for Sub-activities requires clarification
- Enhancement percent might be better replaced with a discrimination between baseline and new capability
- Latent defects would be a good attribute to collect

Future Work

- This research is worthy of a second pass, preferably with new data added to the mix
- Gain further clarity on discrepancies identified between Maintenance Hours and Support Hours and their component hour counts
- Additional data would provide an opportunity to create additional benchmark scenarios
- Application of machine learning techniques such as K-nearest neighbor and neural networks to this dataset

Questions?